A Publication of the American Psychological Association

# The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

Recency and Primacy in Persuasion as a Function of the Timing of Speeches and Measurements ..... 1
*Norman Miller and Donald T. Campbell*

The Interference Theory of Forgetting and the Amnesic Syndrome ..... 10
*George A. Talland*

Social Isolation vs. Interaction With Adults in Relation to Aggression in Preschool Children ..... 17
*Willard W. Hartup and Yayoi Himeno*

Personal Motivations and Self-Conceptions Associated With Effectiveness and Ineffectiveness in Emergent Groups ..... 23
*Paul H. Mussen and Lyman W. Porter*

Regional Differences in Anti-Negro Prejudice ..... 28
*Thomas F. Pettigrew*

Relationships of Measures of Anxiety and Experimental Instructions to Word Association Test Performance ..... 37
*Irwin G. Sarason*

Attribute Criteriality and Sex-Linked Conservatism as Determinants of Psychological Similarity ..... 43
*Michael A. Wallach and Albert J. Caron*

Relationships Between Social Need Strivings and the Development of Heterosexual Affiliations ..... 51
*William J. Meyer*

*Continued on inside front cover*

Volume 59 JULY 1959 Number 1

CONTENTS—*Continued*


Motivation and Language Behavior: A Content Analysis of Suicide Notes . . 58
*Charles E. Osgood and Evelyn G. Walker*

Group Versus Individual Problem-Solving . . . . . 68
*William L. Faust*

Differences in Gradients of Stimulus Generalization as a Function of Psychiatric Disorder . . . . . 73
*Irwin J. Knopf and Robert E. Fagot*

Attractiveness of Group as Function of Self-Esteem and Acceptance by Group . . . . . 77
*James E. Dittes*

Process-Reactive Schizophrenia and Genetic Levels of Perception . . . . . 83
*Harold J. Fine and Carl N. Zimet*

A Method for Determining Types of Self-Esteem . . . . . 87
*Stanley Coopersmith*

Conflict and Repression During Sleep Deprivation . . . . . 95
*Edward J. Murray*

Suggestibility as Dependency Behavior . . . . . 102
*Leonard P. Ullmann and Richard H. Walters*

Perceptual Preferences as a Function of Personality Variables Under Normal and Stressful Conditions . . . . . 108
*Natalia Potash*

Verbal Discrimination of Pleasant and Unpleasant Stimuli Prior to Specific Identification . . . . . 114
*Charles W. Eriksen, Hiroshi Azuma, and Rosalie B. Hicks*

The Effects of a Monetary Reward on the Acquisition of an Imitative Response . . . . . 120
*John T. Lanzetta and Vera T. Kanareff*


CRITIQUE AND NOTES


Phantom Experiences in Mental Defective Amputees . . . . . 128
*Marianne L. Simmel*

Social Perception and the Motivational Structure of Interpersonal Relations . 130
*Milton E. Rosenbaum*

Regression Under Stress to First Learned Behavior . . . . . 134
*Richard P. Barthol and Nani D. Ku*

The Afterimages of "Hallucinated" and "Imagined" Colors . . . . . 136
*Theodore Xenophon Barber*

Salience of Norms and Order of Questionnaire Items: Their Effect on Responses to the Items . . . . . 139
*Charles Y. Nakamura*

Authoritarianism and Reactions to "Sputniks" . . . . . 142
*Walter Mischel and John Schopler*

Contradictory Results Using the Mecholyl Test to Differentiate Process and Reactive Schizophrenia . . . . . 145
*Marvin Zuckerman and Howard J. Gross*

Concerning the Concreteness in Schizophrenia . . . . . 146
*Kurt Goldstein*

Patterns of Parenthood as Theories of Learning . . . . . 148
*Jane Loevinger*

Volume 59 JULY, 1959 Number 1

# The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
*Harvard University*

URIE BRONFENBRENNER
*Cornell University*

ROGER BROWN
*Massachusetts Institute of Technology*

DONALD T. CAMPBELL
*Northwestern University*

LAUNOR F. CARTER
*System Development Corporation*

RICHARD CHRISTIE
*Columbia University*

MORTON DEUTSCH
*Bell Telephone Laboratories*

I. E. FARBER
*State University of Iowa*

DONALD W. FISKE
*University of Chicago*

ROGER W. HEYNS
*University of Michigan*

ROBERT R. HOLT
*New York University*

IRVING L. JANIS
*Yale University*

HAROLD E. JONES
*University of California*

HAROLD H. KELLEY
*University of Minnesota*

LEONARD S. KOGAN
*Community Service Society of New York*

DANIEL J. LEVINSON
*Harvard Medical School*

DAVID C. MCCLELLAND
*Harvard University*

LEO POSTMAN
*University of California*

ELIOT H. RODNICK
*Duke University*

NEVITT SANFORD
*University of California*

SEYMOUR SARASON
*Yale University*

RICHARD L. SOLOMON
*Harvard University*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 832 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

M. BREWSTER SMITH
Department of Psychology
University of California
Berkeley 4, California

---


| ARTHUR C. HOFFMAN | HELEN ORR | BARBARA CUMMINGS |
|---|---|---|
| Managing Editor | Promotion Manager | Editorial Assistant |


Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.


PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices



Copyright 1959 by the American Psychological Association, Inc.

# RECENCY AND PRIMACY IN PERSUASION AS A FUNCTION OF THE TIMING OF SPEECHES AND MEASUREMENTS[1]

NORMAN MILLER AND DONALD T. CAMPBELL

*Northwestern University*

IN A recent volume, Hovland, Campbell, Brock, Luchins, Cohen, McGuire, Janis, Feierabend, and Anderson (1957) have summarized and added to the perplexing literature on the conditions under which the first or the second of two opposing arguments have an advantage due to position per se. Their findings seem to have ruled out any completely general principle of primacy in persuasion, but have specified several sets of conditions under which a primacy effect may be expected. In adding to this growing literature, the present paper pays relatively less attention to primacy, although we do find a primacy effect. Instead, attention is called to a pervasive and probably universal psychological principle generating recency effects, a principle about which psychology already knows enough to predict confidently the temporal conditions of presentation and measurement under which recency should appear most strongly. As a complement, the principle also predicts the temporal conditions under which a primacy effect appears most strongly, if indeed one is present.

Our judgments, our responses, our social perceptions, are a function of some net resultant of the past experiences both recent and remote. To the net resultant of the moment not all past experiences contribute equally: the casual experiences of this morning weigh more heavily than the comparably casual experiences of any single morning one year ago, one month ago, or even yesterday, all other things being equal. Were it not so, unlearning and new learning could hardly take place. But the advantage that this morning's experience now has will dissipate rapidly. By next month, its advantage over yesterday's contribution will be scarcely noticeable. The momentary advantage of the very recent may allow trivial events of this morning to overweigh momentarily more significant learnings of the past, but this momentary advantage will dissipate rapidly, allowing the relative influence of the older learnings to recover spontaneously tomorrow or the next day, if events like this morning's prove to be untypical and do not recur.

These homely truisms are economically summarized in the oldest dependable achievement of the scientific study of learning, the negatively accelerated forgetting curve of Ebbinghaus (1913), and in its implications for the relative strength of competing associations of different ages and strengths. From it, Jost's (1897) second law can be derived—of two associations equally strong at the moment, the older will decay less rapidly. These same considerations also predict spontaneous recovery (Miller & Stevenson, 1936). Hovland, Janis, and Kelly (1953, p. 126), although not making explicit use of the forgetting curve, have provided a parallel analysis of the effect of delays in measurement upon the relative strength of a first and second message:

> Yet another important factor to consider is the time interval between the learning of the second communication and the occasion when remembering is required. In experimental results on simple verbal material one would predict recency effects when the time interval is short. But the passage of time would tend to decrease this recency effect and permit the other factors making for primacy to become relatively stronger. Underwood [1948] has shown that for verbal learning the relative superiority of the second list with respect to retention decreases as the time between learning and testing becomes greater. If the same factors operate in social communication one would expect to find that when the issue is raised quite a while after the communication, recency would be less likely to be operative. Striking data are offered on this point in an experiment by Bateman and Remmers [1941].... The results show that immediately after the two communications the attitudes changed in the direction of the second communication. This is the predicted recency effect. Two months later however, the attitudes had changed in the direction of the first communication. Thus primacy effects occurred with time. While the data are consistent with the analysis suggested here, the experiment cannot be

[1] This paper is based upon a Master of Science thesis (1957) done by the first author under the direction of the second. Additional analyses of the data beyond the original report have been supported in part by funds from the Graduate School, Northwestern University.

considered as presenting conclusive evidence because it did not include a control group and the order of presentation was not rotated.

The present study examines with better controls this effect of delay in measurement, and in addition considers the still stronger factor of the time interval between the first and second communication.

Figure 1 presents the implications of the Ebbinghaus curve as it applies to the relative weighting of two competing communications presented and measured at varying intervals, assuming no interaction of one with the other. In this diagram, the solid line (Line A) represents the contribution of a communication presented first, its strength decaying as time elapses. The two dashed lines (B and B′) represent the contributions of a second opposing communication. In one instance this second communication is presented immediately after the first (Line B), in the other instance, one week after the first (Line B′). Net effects of the two communications in combination are reflected by the vertical distance between Line A and Line B or B′ at any given point in time. The numbered vertical slicings in the diagram represent four possible schedulings of a measure of net effect, designated as Conditions 1–4. Five predictions emerge, stated in terms of the relative magnitude of the recency effect, i.e., the dominance of the second communication in the composite under the four conditions: $3 > 4$, $3 > 1$, $3 > 2$, $1 > 2$, and $4 > 2$. The relative magnitudes of Conditions 1 and 4 cannot be predicted without specific knowledge of the parameters of the curve. The others follow from the simple assumption of similar negatively accelerated decrements for the three curves involved.

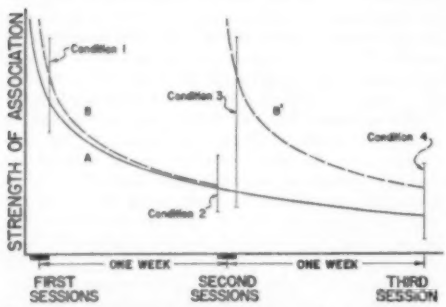


FIG. 1. HYPOTHETICAL FORGETTING CURVES FOR A FIRST AND A SECOND COMPETING PERSUASIVE COMMUNICATION WHEN BOTH ARE PRESENTED DURING A SINGLE SESSION (CURVES A AND B) OR WHEN THE TWO ARE PRESENTED ONE WEEK APART (CURVES A AND B′); THE FOUR CONDITIONS, APPEARING AS VERTICAL SLICINGS, REPRESENT THE TIMING OF MEASUREMENTS

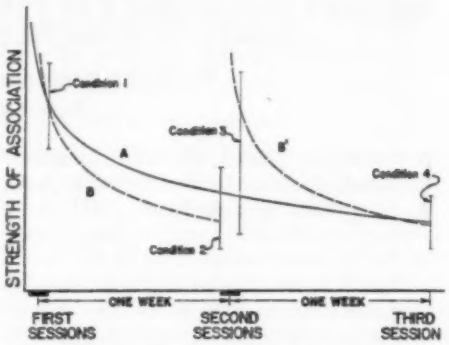


FIG. 2. HYPOTHETICAL FORGETTING CURVES FOR TWO COMPETING COMMUNICATIONS WITH AN ADDED PRIMACY EFFECT; THE PRIMACY EFFECT IS REPRESENTED BY THE HIGHER INITIAL STARTING POINT AND FINAL ASYMPTOTE OF LINE A IN COMPARISON WITH LINES B OR B′

Figure 1 has been drawn assuming equal initial strength for the first and the second messages at the time of their delivery. But the ordering of the test conditions as to their *relative* favorableness to the second message is independent of this assumption of initial equality. These general decremental processes occur no matter how strong any given message happens to be. As can be seen from Figure 1, by assuming this initial equality in the strength of the two communications, though the magnitude of the effects varies, *only* recency effects can be generated. However, in persuasion studies there frequently appears an advantage to the first message, a primacy or prior entry effect, strong enough to show up even under conditions of testing such as Condition 1 in which a weak primacy effect would be covered over by a recency effect. Figure 2 has been drawn to indicate some degree of primacy effect, interpreted as an initially higher level for the first message at its onset than for the second message at its highest and, correspondingly, a higher eventual asymptotic level of strength for the first message than for the

second. Under this assumption, a manifest or net primacy can occur, as in Condition 2. Note that such a primacy effect does *not* occur *instead of* a recency effect but, rather, *in addition* to a recency effect. Thus, attention is drawn to a conceptual distinction between recency effects which are a function of the general rates of decrement with the passage of time, and primacy effects which appear as a function of the higher asymptote eventually resulting from the advantage of prior entry. To predict whether primacy or recency will appear at any single point in time, knowledge of the specific parameters of the curves would of course be essential. But even without specification of parameters, solely upon the basis of assuming similar negatively accelerated decrements, the rank order of the four testing situations in terms of the relative advantage to recency can be specified, and are the same for Fig. 2 as for Fig. 1. However, the absolute magnitudes have been shifted, so that under Condition 2, which is the least advantageous condition for recency in both figures, a strong net primacy effect is predicted in Fig. 2. In general, the more delayed the test period relative to the time gap between the two presentations, the more favorable are conditions for the appearance of a primacy effect.

## METHOD

### *Subjects*

A total of 144 Northwestern freshman and sophomore students from introductory psychology and social science courses served as *Ss*. All *Ss* were volunteers. Some of the *Ss* were acquainted with the experimenter, but these were distributed in approximately equal numbers among the groups serving under different experimental conditions.

### *Communications*

A desirable consideration that has not been met in most two-sided communication studies is that the subject matter or theme of the communications be one about which *Ss* have not already formed opinions or attitudes prior to the experiment. Learning prior to the experiment can eliminate the possibility of controlling or determining the effect of order of learning (Hovland et al., 1953; Lund, 1925). With such the case, primacy of the communication that is experimentally presented first may be a misnomer if the communication is considered in the context of all the communications received by the individual on the particular issue in question. It was therefore considered advantageous to experimentally create attitudes rather than arouse previously learned attitudes. As Weld and Roff (1938), we used a modified courtroom procedure to introduce experimentally pro and con statements about a topic with which *Ss* had no prior experience. A transcript of an actual trial involving a suit for damages supposedly incurred as a result of a defective vaporizer was edited for the purposes of the present experiment.[2] Repetitious sections of the testimony were condensed and the names of the principals were changed. The sequence of the proceedings was rearranged so that all material for the plaintiff appeared in one consecutive block, and all material for the defendant appeared in another block. The testimony of witnesses for the plaintiff, the cross-examination by the plaintiff's lawyer of the witnesses for the defense, and the opening and closing speeches of the plaintiff's lawyer, will be called the "Pro" communication, the testimony of witnesses for the defense, etc. the "Con." These two communications were approximately equal, both in length and in the number of points to their arguments. Slight differences in persuasive strength of the two communications were present but are unimportant because of the counterbalanced design employed. The experimental communications were recorded, with different persons reading the parts of the different characters, lawyers, witnesses, etc. Each took approximately 45 minutes to run.

### *Response Measures*

Four response measures were used. A graphic scale with nine described points (e.g., defendant is somewhat more responsible) enabled each *S* to express the degree to which he thought either the plaintiff or the defendant was responsible for the accident, the midpoint of 5 representing equal responsibility. Higher values indicated plaintiff's responsibility, i.e., persuasion to the "Con" point of view. This scale served as the direct attitude measure. It was accompanied by a scale on which the *S* estimated how certain he was of his response, with possible values ranging from zero to 100%.

A multiple choice test was constructed to determine the amount of factual information that could be recalled from the two communications. The test contained a total of 28 four-choice items, 14 based on the material from the Pro communication and 14 based on the material from the Con communication.

The one other response measure used was included in an effort to get at differences in acceptance. *Ss* were requested to list those parts of the two communications that they could not believe or accept. This response measure was preceded by a statement to the effect that no person believes everything he is told even if it is supposedly factual. Then, a list of the names of the persons who testified at the trial (a brief identifying description following each name) was supplied to each *S* in order to facilitate recall and identification of unacceptable statements. The response measures collectively took approximately 30 minutes.

### *Procedure*

The two spacings of the first and second communication (immediate and one week apart), and, for each, the

[2] The authors are indebted to Fred L. Strodtbeck of the University of Chicago for this material.

two spacings of the measurement efforts (immediately after the second and one week after the second) generate the four experimental conditions indicated in Figs. 1 and 2. Replicating these four conditions with the Pro-Con order and the Con-Pro order generates the eight basic experimental groups. Each of these groups contained 18 *Ss*, all presented the communications and tested together. All groups met in the same classroom at about the same time of day. *Ss* sat one seat apart from each other, and adjacent *Ss* had booklets in which the test materials were assembled in different orders. Through these precautions it was hoped to minimize the "neighbor effect" or "social diffusion," well-known reducers of the variability of the test scores of college students. Each group was measured but once in a Posttest Only Design (Campbell, 1957). For Condition 1, the single session took approximately 2½ hrs. with 5-minute breaks following both communications.

To make the findings on the various measures independent of the order in which the students filled them out, the three instruments (attitude and certainty being treated as one) were stapled together in the six possible orders, and three of each of these assemblies employed in each group of 18. Extra *Ss* were run in each group to allow for the inevitable loss through failure to keep appointments one week apart. Surplus *Ss* were discarded to leave 3 *Ss* each in the 48 cells of the total design. Since the order of the response measures turned out to have no effect in any of several analyses (Miller, 1957) it has been omitted in this report.

## RESULTS

*Attitude.* In Table 1 are presented the mean attitude scores for each of the eight groups. Using Table 1 as a reference, two types of comparisons can be designated. The vertical comparisons within each of the two columns are appropriate to testing the predictions about the relative strength of the second message. For this purpose, each column represents an independent experiment of four groups. It is for these comparisons that the Ebbinghaus curve predictions are made. Table 2 and the discussion of it below treat these results. The predictions in Table 2, and the columns of means of Table 1 taken individually, are silent on whether or not for any given condition the manifest effect found is one of primacy or recency. The predictions merely indicate under which condition a primacy effect, if present, should appear most strongly (i.e., Condition 2) and under which a recency effect, if present, should appear most strongly (i.e., Condition 3), etc. For any one mean to be directly interpretable in terms of primacy and recency, one would need proof that the Pro and Con messages were exactly equal in persuasive power and that the assigned midpoint of 5 was the effective neutral point of judgment in absence of persuasive communications. But if one makes use of the comparison of the Pro-Con and the Con-Pro orders, i.e., makes the horizontal comparisons of Table 1, the presence of manifest net primacy or recency effects can be examined, because the groups being compared have heard identical materials under identical spacings, and differ only in the order with which these have been heard. In Table 1 (and Tables 4 and 7), the higher the scores the more the Con material has been effective. Thus a comparison of the two Condition 2 groups shows that the Con material was weakest when occurring second, strongest when coming first, indicating a primacy effect, significant at the .01 level (using a two-tailed test, since direction was not predicted). Condition 3 shows a recency effect, significant at the .02 level. The results conform to the pattern predicted by Fig. 2, in which a general primacy and recency effect coexist, the recency effect dominating under Condition 3, the condition predicted as most favorable to recency; the primacy effect dominates under Condition 2, predicted as least favorable to recency, and therefore as the place where a primacy effect would appear most strongly, if present. The two effects cancel each other in the two conditions of predicted intermediacy.

While the $t$ tests reported in Table 1 and Table 2 provide more specific information, a double classification analysis of variance offers a more general test of the main hypothesis that the temporal spacing of communications and measurements determines the relative effec-

TABLE 1
MEAN ATTITUDE SCORES

| Condition | Pro-Con Presentation | Con-Pro Presentation | Diff. | $t$ | $p$** | Direction |
|---|---|---|---|---|---|---|
| 1($X_1X_2O$)* | 5.94 | 5.88 | .06 | | | |
| 2($X_1X_2$—$O$) | 4.50 | 6.61 | −2.11 | 2.81 | .01 | Primacy |
| 3($X_1$—$X_2O$) | 6.00 | 4.33 | 1.67 | 2.46 | .02 | Recency |
| 4($X_1$—$X_2$—$O$) | 5.47 | 5.58 | −.11 | | | |

* The shorthand characterization of the conditions employs a standardized symbolization of experimental treatments (Campbell, 1957), in which $X_1$ and $X_2$ represent the experimental stimuli, $O$ the process of observation and measurement, and spacing from left to right the time dimension.

** Two-tailed, inasmuch as direction of difference not predicted.

tiveness of the communications in the composite. However, with a counterbalanced design, this hypothesis must be stated as an interaction rather than a main effect; the shifts in attitude occurring over conditions with the Pro-Con presentation should be analogous, but reciprocal or in opposite direction to those occurring with the Con-Pro order. In a 2 × 4 analysis of variance of the attitude scores, the $F$ ratio for the interaction between Conditions and Order of Presentation is the only one that obtains significance. With 1 and 3 degrees of freedom, this $F$ ratio of 4.06 is significant beyond the .01 level.

Table 2 presents, separately, for the Pro-Con and Con-Pro orders, the comparisons between conditions in terms of their predicted relative favorableness to recency. The positive signs indicate that the direction of difference was as predicted. (Note that for the Con-Pro groups, greater recency effects are indicated by lower scores.) All differences are in the direction predicted. This score of 10 correct out of 10 predictions looks highly significant by a simple sign test, although since the same values have been re-used, independence of sampling error from comparison to comparison cannot be claimed. Table 2 also presents the $t$ ratios and $p$ values for the predicted comparisons of means and a compound $p$ value (Baker, 1952) for the Pro-Con and the Con-Pro groups treated jointly. The most clearly significant difference occurs for the comparison for which the largest contrast was expected, i.e., between

TABLE 2
COMPARISON OF MEANS FOR ATTITUDE SCORES[a]

| Prediction for Conditions | Pro-Con Presentation | | | Con-Pro Presentation | | | Combined |
|---|---|---|---|---|---|---|---|
| | Diff | $t$ | $p$ | Diff. | $t$ | $p$ | $p$ |
| 1 > 2 | +1.44 | 1.79 | .04 | +.73 | 1.00 | .16 | <.05 |
| 3 > 1 | +.06 | .08 | .47 | +1.55 | 2.03 | .025 | <.07 |
| 3 > 2 | +1.50 | 2.14 | .02 | +2.28 | 3.12 | .005 | <.001 |
| 3 > 4 | +.53 | .78 | .22 | +1.25 | 1.56 | .06 | <.08 |
| 4 > 2 | +.97 | 1.21 | .11 | +1.03 | 1.35 | .09 | <.06 |
| (1, 4) | .47 | .59 | .56 | .30 | .38 | .70 | |

[a] The Diff. column values are differences between the group means presented in Table 1, with positive signs applied to indicate differences in the predicted direction as favoring recency, thus reversing the natural subtraction signs for the Con-Pro groups for which favoring recency implies lower scores. No prediction was made for Conditions 1 vs. 4, and two-tailed $p$ values are therefore presented for this comparison, while one-tailed values are used for the others.

TABLE 3
MEAN NUMBER OF ITEMS CORRECT ON RECALL

| Condition | Pro-Con Presentation | | Con-Pro Presentation | |
|---|---|---|---|---|
| | Pro | Con | Pro | Con |
| 1($X_1X_2O$) | 10.00 | 9.44 | 10.72 | 9.83 |
| 2($X_1X_2$—$O$) | 9.28 | 9.06 | 8.94 | 8.33 |
| 3($X_1$—$X_2O$) | 7.31 | 9.69 | 9.94 | 6.67 |
| 4($X_1$—$X_2$—$O$) | 7.50 | 8.50 | 7.78 | 6.89 |

Conditions 3 and 2. The other differences approach significance in the predicted directions. All in all, results seem strongly confirmatory of the predictions.

*Recall of Information.* Table 3 presents the recall data in terms of the mean number of Pro and Con items correctly answered in each condition. To state recall in terms of a net weighting, each $S$ was given a score of Con items right minus Pro items right. The means of these net recall scores are presented in Table 4. The comparisons made in Table 4 parallel those in Table 1. In an analysis of variance of the net recall scores, a significant interaction obtains between Conditions and Order of Presentation. With 1 and 3 degrees of freedom, the $F$ ratio of 21.46 for the interaction is significant well beyond the .001 level. All horizontal comparisons show a net recency effect (conforming more to the model of Fig. 1 than Fig. 2), with the recency effect most highly significant for Condition 3, as expected.

Table 5 is constructed parallel to Table 2. Of the ten predicted differences, nine are confirmed for sign. The three predictions involving Condition 3 are all very highly significant. The comparison between Conditions 1 and 2 fails of significance in either the Pro-Con or Con-Pro order. Considering the small number of items in the test, and their employment as a difference score (which should further increase their random variability), the results seem in general as confirmatory as one could expect. In both columns, two independent replications, the gap between Condition 3 and any other is most dramatically confirmed, as would be expected from Fig. 1.

*Nonacceptance.* The mean number of Pro and Con arguments recalled as unbelieved under the eight conditions is shown in Table 6. As a net effect score, the number of unaccept-

TABLE 4
MEAN NET RECALL SCORES

| Condition | Pro-Con Presentation | Con-Pro Presentation | Diff. | $t$ | $p$* | Direction |
|---|---|---|---|---|---|---|
| $1(X_1X_2O)$ | −.56 | −.89 | +.33 | .39 | | |
| $2(X_1X_2—O)$ | −.22 | −.61 | +.39 | .55 | | |
| $3(X_1—X_2O)$ | 2.38 | −3.27 | +5.65 | 7.90 | .001 | Recency |
| $4(X_1—X_2—O)$ | 1.00 | −.89 | +1.89 | 2.62 | .02 | Recency |

* Two-tailed.

TABLE 5
COMPARISON OF MEAN RECALL DIFFERENCE SCORES

| Prediction for Conditions | Pro-Con Presentation | | | Con-Pro Presentation | | | Combined |
|---|---|---|---|---|---|---|---|
| | Diff.* | $t$ | $p$ | Diff. | $t$ | $p$ | $p$ |
| 1 > 2 | −.34 | .39 | .64 | +.28 | .40 | .36 | > .20 |
| 3 > 1 | +2.94 | 3.42 | .0004 | +2.39 | 3.21 | .001 | < .0001 |
| 3 > 2 | +2.60 | 3.80 | .0001 | +2.67 | 3.54 | .0002 | < .0001 |
| 3 > 4 | +1.38 | 1.76 | .04 | +2.39 | 3.52 | .0003 | < .0001 |
| 4 > 2 | +1.22 | 1.54 | .065 | +.28 | .45 | .34 | < .10 |
| (1, 4) | 1.56 | 1.67 | .10 | .00 | .00 | .50 | |

* See the footnote to Table 2.

able Con items was subtracted from the number of unacceptable Pro items. The means of these net scores for each group are presented in Table 7. For this measure, the interaction $F$ ratio for Order × Conditions was insignificant. The Pro-Con vs. Con-Pro differences are predominantly in the primacy direction, although not of clear statistical significance. The most significant difference is in the primacy direction under Condition 2, predicted most favorable to primacy. The only difference in the recency direction is in Condition 3, predicted most favorable to recency. Table 8 presents the differences among the Pro-Con conditions, and among the Con-Pro conditions, arranged in terms of the predictions, although it is not clear that the predictions are appropriate to this measure. For the Pro-Con groups, all differences are trivial, and the predicted direction of difference is found in only two of the five pairs. The Con-Pro groups offer four out of five confirmations, all those that approach significance being in the predicted direction.

In retrospect, it can be seen that our particular measure of nonacceptance is far from ideal. A more interpretable procedure would have asked for recall of testimony found particularly convincing in addition to testimony disbelieved. With only the recall of unacceptable material, the influence of credibility is mixed with the influence of greater readiness of recall, with opposing effects. For an item to be mentioned as unbelievable it has to be remembered. But mention of second message material as disbelieved is scored as a credibility advantage to the first. If, for the reason of greater recency, second message material is more easily remembered and therefore mentioned more often, the net effect gets scored as a primacy advantage for acceptance and credibility purposes.

*Other Analyses.* To assist in the interpretation of the processes that would allow a primacy effect for the attitude measure without a primacy effect for recall, correlation coefficients were computed among the three main variables, separately for each of the eight groups. The average of these values are: Attitude and Recall −.10, Attitude and Nonacceptance .43 ($p < .01$), Recall and Nonacceptance −.04. For the last two, the correlations for the individual groups are adequately represented by these averages. However, as a provocative item for contemplation if confirmed in future research, it can be noted that

TABLE 6
MEAN NUMBER OF STATEMENTS FOUND UNACCEPTABLE

| Condition | Pro-Con Presentation | | Con-Pro Presentation | |
|---|---|---|---|---|
| | Pro | Con | Pro | Con |
| $1(X_1X_2O)$ | 1.28 | .56 | 2.27 | .33 |
| $2(X_1X_2—O)$ | 1.39 | .78 | 2.27 | .44 |
| $3(X_1—X_2O)$ | 1.31 | .81 | 1.39 | .95 |
| $4(X_1—X_2—O)$ | 1.33 | .39 | 1.33 | .39 |

TABLE 7
MEAN NET NONACCEPTANCE SCORES

| Conditions | Pro-Con Presentation | Con-Pro Presentation | Diff. | $t$ | $p$* |
|---|---|---|---|---|---|
| $1(X_1X_2O)$ | .72 | 1.94 | −1.22 | 1.66 | .10 |
| $2(X_1X_2—O)$ | .61 | 1.83 | −1.22 | 1.91 | .06 |
| $3(X_1—X_2O)$ | .50 | .44 | .06 | .10 | .92 |
| $4(X_1—X_2—O)$ | .94 | .94 | .00 | .00 | .50 |

* Two-tailed.

TABLE 8
COMPARISON OF MEAN NET NONACCEPTANCE SCORES[a]

| Comparison | Pro-Con Presentation | | | Con-Pro Presentation | | | Combined |
|---|---|---|---|---|---|---|---|
| | Diff. | $t$ | $p$ | Diff. | $t$ | $p$ | $p$ |
| 1 > 2 | +.11 | .15 | .44 | −.11 | .17 | .57 | >.20 |
| 3 > 1 | −.22 | .31 | .64 | +1.50 | 2.27 | .012 | <.05 |
| 3 > 2 | −.11 | .17 | .57 | +1.39 | 2.21 | .014 | <.05 |
| 3 > 4 | −.44 | .98 | .84 | +.50 | .83 | .21 | >.20 |
| 4 > 2 | +.33 | .64 | .26 | +.89 | 1.53 | .06 | <.10 |
| (1, 4) | .22 | .38 | .65 | 1.00 | 1.63 | .10 | |

[a] See footnote for Table 2.

the correlations between Attitude and Recall seem to show systematic variance from substantial positive correlation under Condition 1 (.49 and .27) to substantial negative correlation under Condition 4 (−.51 and −.40).

As reported more fully elsewhere (Miller, 1957), certainty scores proved unrelated to the experimental conditions. Similarly, certainty of attitude did not correlate with amount of recalled information, nor did the usual curvilinear correlation between certainty and attitude-extremity appear. While the information test was not deliberately designed for the purpose, it was possible to order the wrong alternatives in the degree to which they implied conclusions favorable to the Pro or the Con argument, and to give each *S* a score indicating the preponderant directionality of his errors. This indirect attitude test (Campbell, 1950) correlated but .08 with the direct attitude score and showed no systematic relationship to the experimental conditions.

## DISCUSSION

*Recency.* The clear-cut results for the attitude and recall measures clearly support the application of the Ebbinghaus curves to the field of persuasion. Nor would there seem to be any rival theory that could so economically predict the complex set of results. The ubiquity of the negatively accelerated forgetting curve in learning studies further supports the possibility of a thoroughly general recency principle, with the strength of recency being maximal for a long delay between first and second communication coupled with immediate measurement after the second communication, the strength of recency being minimal when the two presentations are contiguous and the testing delayed. One study does not make a general principle, and the area of persuasion has been particularly frustrating to such attempts, but the antecedent probabilities plus this clear-cut experimental confirmation give promise. The recognition that a recency effect (such as the Ebbinghaus curves predict) can be a concomitant of, rather than an excluding alternative to, a primacy effect makes the postulation of this general principle not in the least incompatible with any of the previous literature on order effects. Indeed, the only other study containing both a question of order of arguments and measurement at various removes in time confirms some of the predictions of this principle, although that study is by today's standards so lacking in controls as to be uninterpretable (Bateman & Remmers, 1941).

Ebbinghaus' curve is simple and perhaps obvious, yet, as indicated, the simultaneous application of several such curves generates the nonobvious Jost (1897) law number 2, and even spontaneous recovery (Miller & Stevenson, 1936). Considering the length of time that the principle has been with us and its potential fertility, it is surprising that it has been so little used. One of the historical reasons is that with the observations of lack of forgetting during sleep and more rapid forgetting with similar materials during the interposed period, forgetting became something to be explained, and therefore attention was distracted from its use as an explanation. But even accepting an interference theory of forgetting, and even employing a psychological time scale (perhaps average interference rate during waking hours) the negatively accelerated curve has explanatory power and can be accepted at a molar level as a given for social psychology.

One of the most general findings in persuasion studies, involving repeated measurements over a long period of time, is that the change induced by persuasion dissipates with time. Similarly in everyday observations, successful remedial efforts and conversions are often followed by recidivism and backsliding. Usually such phenomena are interpreted as due to the competing effects of other communications occurring subsequent to the persuasive message, but under the aegis of Ebbinghaus'

curves, one can interpret this decrement as a literal backsliding, due to the increasingly effective competition of the great bulk of prepersuasion communications, as time since the persuasive effort increases. On the other hand, when one considers the volumes of communications that *S*s have received prior to an experimental session, it becomes far from obvious that a 10-minute speech on a familiar topic could have any appreciable effect. Yet the most general finding in both published studies and in unpublished ones is that almost any speech will have an effect in its intended direction if the effect is measured immediately. If this fact needs explanation, as we think it does, it is found in the Ebbinghaus curves, which predict an extreme recency advantage under this condition. By the same token, the advantage rapidly dissipates. And if the speech is on a familiar topic, then even if no further communications were to be received, by next year the ten-minute speech will indeed be but a drop in the bucket.

*Primacy Effects Upon Attitude Measures.* Paradoxically enough, the postulation of this general recency principle improves the case for a general primacy principle. The experimental arena employed in primacy-recency studies so far has been most similar to our Condition 1, immediate testing after contiguous presentations. This is a condition that is not optimal to the manifestation of either primacy or recency. In the present study they cancel each other out under this condition. The inconsistent evidence obtained at this point by the half-dozen studies which Hovland et al. (1957) review and report thus could be interpreted, if one wished, as consistent with an always present primacy effect. The strength of this primacy effect would necessarily vary with the experimental conditions particular to each of the different studies and would frequently be masked by the recency effects also present. Testing this assumption would involve, in part, studies with sufficiently long delayed measurements to explore that end of the dimension represented by our Condition 2. It is not the purpose of this paper to argue the case for a general primacy effect but merely to point out that the evidence to date does not rule one out.

Would such a general primacy effect include learning phenomena as well as attitude phenomena? Here, again, test conditions such as Condition 2 are very rare. The trends in the available long delay studies favor primacy (Underwood, 1948), significantly so in one instance (Briggs, 1954). But in these studies, the first list learned was presented for more trials than the second, and none have employed as long a delay as one week. In our own recall data, no slightest hint of a net manifest or phenotypic primacy effect occurs. It seems clear that in the present study the primacy effect in the attitude measure is at very least much stronger than whatever hidden primacy effect might be latent in the recall data. The primacy effect for attitudes is thus not to be explained in terms of a primacy effect upon learning or memory of evidence or argument. Similarly, it cannot be explained in terms of such factors as set, attention, or reinforcement (Hovland et al., 1957), which assume that the second communication is less well listened to, less well assimilated content-wise, or subjected to stronger interference effects and, as a result, less well remembered. Our recall data are contrary to any such assumption. The primacy effect in attitude change seems related in the present instance to acceptance factors, as Hovland anticipated it might be (Hovland et al., 1953, p. 126). In this interpretation, the evidence on nonacceptance, rough though it is, mediates between the attitude and recall data. Coming first gives a statement no greater probability of being remembered, but does give it greater probability of being believed. We have a general tendency to find one side of an argument persuasive, providing we have not heard the other. Hearing it after we have heard the other, we are apt to be more critical and skeptical.

## SUMMARY

The order in which opposing arguments were presented (Pro-Con and Con-Pro), the time interval between them (none and one week), and the time of testing (immediately after the second argument or one week after) were varied for eight groups, each tested once. For the attitude measure, a significant recency effect was found under the conditions most favorable to recency as predicted from the application of Ebbinghaus decay curves. A significant pri-

macy effect was found under the conditions predicted to be least favorable to recency. For recall of information, only recency effects were found, these being strongest under conditions predicted to be most favorable.

## REFERENCES


BAKER, P. C. Combining tests of significance in cross validation. *Educ. psychol. Measmt.*, 1952, **12**, 300–306.

BATEMAN, R. M., & REMMERS, H. H. A study of the shifting attitude of high school students when subjected to favorable and unfavorable propaganda. *J. soc. Psychol.*, 1941, **13**, 395–406.

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.

CAMPBELL, D. T. Factors relevant to the validity of experiments in social settings. *Psychol. Bull.*, 1957, **54**, 297–312.

EBBINGHAUS, H. *Memory*. H. A. ROGER & C. E. BUSSENIUS (Trans.) New York: Teachers College Columbia Univer., 1913. (Original, *Über das Gedächtnis*, Leipzig, 1885).

HOVLAND, C. I., JANIS, I. L., & KELLY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.

HOVLAND, C. I., MANDELL, W., CAMPBELL, ENID H., BROCK, T., LUCHINS, A. S., COHEN, A. R., MCGUIRE, W. J., JANIS, I. L., FEIERABEND, ROSALIND L., & ANDERSON, N. H. *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.

JOST, A. Die Assoziationsfestigkeit in ihrer Abhangigkeit von der Verteilung der Wiederholungen. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 1897, **14**, 436–472.

LUND, F. H. The psychology of belief: IV. The law of primacy in persuasion. *J. abnorm. soc. Psychol.*, 1925, **20**, 183–191.

MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.

MILLER, N. Primacy versus recency: The changing relative effectiveness of two opposing communications with the passage of time. Unpublished master's thesis, Northwestern Univer., 1957.

UNDERWOOD, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38.

WELD, H. P., & ROFF, M. A study in the formation of opinion based upon legal evidence. *Amer. J. Psychol.*, 1938, **51**, 609–628.


Received April 10, 1958.

# THE INTERFERENCE THEORY OF FORGETTING AND THE AMNESIC SYNDROME[1]

GEORGE A. TALLAND

*Harvard Medical School and Massachusetts General Hospital*

DEFECTIVE retention is one of the characteristic symptoms of Korsakoff's psychosis, a disease in which the amnesic disturbance is disproportionately severe as compared with other cognitive disorders. Several students (Burger-Prinz & Kaila, 1930; Lidz, 1942) of this syndrome resorted to the interference theory of forgetting to account for their observations, arguing that the memory disturbance is so marked because later impressions obliterate preceding ones. Apart from their relatively efficient performances on digit span tests (Wechsler, 1917; Zangwill, 1942–43), Korsakoff patients furnish little support for this view. In fact, if there is one instance in which forgetting cannot be attributed to the interference of new learning, it is that of the Korsakoff patient whose memory is by no means consistently better for the most recent event or information. Van der Horst (1932) reported that his patients, while able to carry out three consecutive commissions, sometimes also included among these one they received earlier. The same observation was made on our patients, who often followed an earlier instruction in the place of one more recently given. Similar errors occurred in recall, e.g., when a patient was asked first to read a few lines of poetry, then to cut out a circle marked on a piece of paper, and last to write down the date of the day. When, on completion of the third task, the patient was asked to recall the last three activities he had been engaged on, he might list them correctly, but he was quite as likely to mention only one or two, and to include a couple of others he had actually been doing somewhat earlier. If this was pointed out to the patient, he might still not remember correctly his last three activities, even though the poem, the circle he cut out, and his writing were all spread out in front of him, while no visible evidence of his earlier performances remained within his sight.

[1] This work was supported in part by a Research Grant, M-767 (c) from the National Institute of Mental Health, United States Public Health Service.

From the obliteration theory of defective retention, one would predict that the interference effect of an activity intervening between learning and relearning should be more pronounced in Korsakoff patients than in suitable control *S*s. Our observations, on the contrary, suggested that this effect is entirely absent or negligible in the amnesic syndrome. Therefore, the prediction in the present study was that Korsakoff patients would not show such effects of retroactive interference as can be detected in the normal processes of recall. In other words, the hypothesis tested was that the effects of interference are significantly less in Korsokoff cases than in suitably chosen controls. Predictions derived from this assertion involved no assumption about the particular processes which might account for the amnesic symptoms, i.e., whether the defect occurred in registration, in recall, or in both. Rather, a conclusion drawn from clinical observations was subjected to experimental testing with a view to clearing the ground of theory from a manifestly implausible alternative hypothesis.

## METHOD

The experimental technique employed was the time honored device of memorizing nonsense syllables. However, since meaningless and therefore entirely novel tasks create special difficulties for Korsakoff patients, the experiment was duplicated with meaningful words. Because of the very limited learning capacity of these patients (Gregor, 1909; Wechsler, 1917), the technique selected was a modified saving method, i.e., *S*s were not required to attain a criterion of learning. They were given lists too long for accurate recall and were expected to retain as many items as they could. In Experiment I, recall was tested by unaided reproduction; in Experiment II, by recognition. While immediate recall of nonsense syllables provided a measure of learning in the sense of acquisition, retention could be tested either by delayed recall or by the increment in acquisition after relearning the

same list. In the light of some pilot studies, the latter measure of retention was chosen, and the experiment followed the familiar pattern of testing retroactive interference, retention, and proactive interference, in that order, for nonsense and meaningful material alike. Experiment II, which followed after an interval of 20 months, was restricted to the Korsakoff group, and did not include a test of proactive interference.

*Subjects.* The experimental sample consisted of 14 women and 6 men, patients at the Boston State Hospital, all of whom had a history of alcoholism and were diagnosed by neurologists as having Korsakoff's psychosis in its chronic phase. They had been hospitalized for periods ranging from one to twelve years. Two members of this group were not available for the test of proaction and for the recognition experiment. The control sample for the major experiment was drawn from two sources, 15 outpatients of the Alcoholic Clinic of the Massachusetts General Hospital, and 8 patients who had been on the neurological wards of the Lemuel Shattuck Hospital for at least six months. The former shared with the experimental group a history of alcoholism but, like them, had no alcohol in their system at the time of testing. The neurological patients had no record of alcoholism and were not known to suffer from brain damage other than that related to the normal aging process; they were chosen to match the experimental group for such effects as prolonged hospitalization might have on their cognitive functioning. For the test of proactive interference, the control group consisted of only 12 outpatients. In age the experimental group ranged from 31 to 64 (mean 51.8, *SD* 8.7), control group I from 37 to 60 (mean 45.2; *SD* 7.5), control group II from 36 to 60 (mean 48.8; *SD* 8.5). Judged by the combined subtests of comprehension, vocabulary, and similarities of the Wechsler-Bellevue Intelligence Scale (1944), the mean IQ of the outpatient group was 116, while both hospitalized samples averaged 103.

*Material.* Both nonsense and meaningful words consisted of three letters, always a vowel between two different consonants, e.g., PEZ or SUN. Similarities in print or sound and associations by meaning or rhyme were avoided as far as possible within each list. Experimental *Ss* were tested on lists of 10 items; controls, however, were mostly tested on lists of 15 nonsense syllables and of 20 meaningful words. The reason for this difference was that lists of 10 did not allow for gain on second presentation, since many control *Ss* could memorize that number of words on first learning. Experimental *Ss*, on the other hand, could in no instance match this achievement, and tended to perform markedly less well if given 15 or 20 rather than 10. Since, for the purpose of testing the main hypothesis of the study, each group served as its own control, this dissimilarity in the experimental material could not contaminate the results. Whether a control *S* was given the longer or shorter list was decided by his performance on a pilot list of 10 nonsense syllables; if he correctly recalled seven or more after three minutes' learning, he was given the longer list in each test.

The test words were printed in ¾-in. type, arrayed vertically on a white card 8 by 11 inches. There were two matched test lists of 15 nonsense syllables each, and likewise two lists of 20 meaningful words. The shorter lists consisted of the first 10 items of each. The nonsense lists were so compiled that the five vowels occurred the same number of times, and each syllable had an association value of 40% or less, following Glaze (1928). The parallel lists of meaningful syllables were equated for familiarity (Thorndike & Lorge, 1944). Twenty additional syllables, nonsense and meaningful, were used in the recognition test. Each of these was printed on a separate card, and followed the pattern of consonant–vowel–consonant.

*Procedure.* In tests by recall, *E* told *S* that he would be given a list of words printed on a card, that his task was to remember accurately as many of these as he could, and that he would have three minutes to learn these words, after which he would report as many as he could remember in any order he liked. Care was taken that *S* clearly understood his task. If *S* raised the question whether nonsense syllables should be pronounced or spelled, *E* suggested that they be pronounced, but spelled responses were accepted; in case of doubtful pronunciation, *S* was asked to spell the syllable. When the three-minute period expired, *E* requested that *S* return the list, but test of recall started only after a delay of 30 seconds. No time limit was set for this, and it usually took one to two minutes. In tests of interference, *S* was then given the alternative list, and the same procedure followed; after the second test of recall, *S* received the first list (testing retroaction) or the second list (testing proactive interference) for relearning. Only one list was used for testing retention, and after first recall *S* was allowed five minutes before relearning it, during which he was not occupied with other tasks.

*Ss* were first tested for retroactive interference, alternately beginning their work with List A or B, the other being used for the intermediate task. An *S* whose first test used the lists in order A–B–A would on the following occasion be given List B for learning and relearning, and on the last test he would again start with B and have A for learning and relearning. Intervals between two tests were two weeks or longer. The parallel tests with nonsense syllables and with meaningful words were conducted within one session. *E* recorded all responses, and if asked for confirmation or for other reasons thought this was necessary, nodded affirmatively, whether *S*'s response was correct or not. Repeated presentations of the same syllables were not recorded.

The same procedure was followed in the tests of recognition, except that each of the four experiments was conducted on a different occasion, retention and retroactive interference being tested first with nonsense and subsequently with meaningful material. Learning periods were limited to three minutes, and testing followed a 30-sec. pause, when *S* was presented with 30 cards, spread out on a table, consisting entirely of either nonsense or meaningful words. They included the 10 items of the test list and another 20 syllables, none

of which was drawn from any test list. *S*s were not restricted in the number of their choices or in the time they took to show their recognition of the syllables. This they did by shifting the cards to the side of the table so that they could see how many they had already selected. In every instance they completed their task within two minutes.

## RESULTS

The number of syllables or words correctly recalled or recognized in each test constituted an *S*'s individual score. Statistical analysis consisted of tests of differences between mean group scores. Since the two control groups consistently deviated from the experimental sample in the same direction, and since their mean scores were invariably much closer to each other than to those of the Korsakoff patients, there was no reason for treating them separately and thus comparing small samples.

In neither group of *S*s and for neither type of test material—nonsense or meaningful—were there significant differences in recall scores between the alternate lists. Mean scores and standard deviations on each test are presented separately for the two samples in Table 1. The control group's scores are consistently higher, and significantly so at the .01 level. Table 2 shows the extent of retention on each test, as measured by the increment of recall after relearning. While all but a few control *S*s improved their recall score on second learning, the Korsakoff *S*s were as likely to score lower on relearning the same list as to register some gain, and any gains were negligible. Comparing groups, $t$ ratios calculated on means of individual increments show that the controls retained some meaningful words if there was neither retroactive nor proactive interference, and some nonsense syllables even though there was interference of either type. Under none of the experimental conditions did the Korsakoff group show this effect of learning. Retroactive and proactive interference was further tested by $t$ ratio for correlated means, in this instance the scores consisting of differences between the same *S*'s gain in relearning without and with interference. The control group's relearning score was decisively reduced by interpolation of an interfering task for nonsense and meaningful material alike ($t = 4.16$ and 6.14, both $p < .01$), but gave no evidence of proactive interference. The Korsakoff group showed no significant gain in relearning either type of material when there was no interference, nor any loss following interference.

Experimental and control groups alike remembered more meaningful words than nonsense syllables under all the test conditions. Measured by additional items recalled, the margin of meaningful over nonsense syllables was wider in the control group, with a mean of 3.73 as against the Korsakoff patients' 2.45 across all tests. Proportionately, however, the experimental group benefited more, with a ratio of 1.95 as against the controls' 1.57. Both groups reported more incorrect nonsense syllables than meaningful words, the means being 1.84 and .84 for the experimental, 1.25 and

TABLE 1
MEANS AND STANDARD DEVIATIONS OF RECALL SCORES

| Test | Retroact. Interfer. | | | No Interference | | Proactive Interfer. | | |
|---|---|---|---|---|---|---|---|---|
| | I | II | III | I | II | I | II | III |
| Korsakoff Group | ($N = 20$) | | | ($N = 20$) | | ($N = 18$) | | |
| Nonsense syllables | | | | | | | | |
| Mean score | 2.85 | 2.15 | 2.75 | 2.45 | 2.70 | 2.50 | 2.50 | 2.78 |
| *SD* | 2.18 | 1.71 | 1.21 | 1.38 | 1.56 | 1.20 | 1.62 | 1.83 |
| Meaningful words | | | | | | | | |
| Mean score | 4.60 | 4.90 | 5.45 | 5.25 | 5.10 | 4.61 | 5.17 | 5.17 |
| *SD* | 1.57 | 2.27 | 2.34 | 1.86 | 1.72 | 2.23 | 1.85 | 2.28 |
| Control group | ($N = 23$) | | | ($N = 23$) | | ($N = 12$) | | |
| Nonsense syllables | | | | | | | | |
| Mean score | 5.70 | 5.26 | 6.83 | 5.65 | 7.39 | 7.92 | 6.75 | 8.75 |
| *SD* | 2.82 | 2.32 | 2.74 | 2.74 | 3.04 | 3.03 | 3.22 | 3.47 |
| Meaningful words | | | | | | | | |
| Mean score | 8.96 | 8.83 | 9.52 | 10.26 | 12.26 | 10.67 | 10.83 | 12.50 |
| *SD* | 3.40 | 3.78 | 3.36 | 2.85 | 3.68 | 4.19 | 2.82 | 4.19 |

.32 for the control *S*s. Taking the proportion of incorrect to total responses as an error score, the means are .47 and .15 for the former, .16 and .03 for the latter; the mean difference between error score for nonsense and for meaningful material is significant ($t = 7.4$, $p < .01$) in the experimental group, but not ($t = 1.78$) in the control. Using the Mann-Whitney test (1947) because of heterogeneous variances, the error scores of the experimental group were found to be significantly larger ($p < .01$) than those of the control group for nonsense and meaningful material alike. It should be noted, however, that on more than half the tests with meaningful words and about one sixth of the tests with nonsense syllables, the Korsakoff group did not make a single incorrect response, and that 12 of the 20 experimental *S*s averaged less than one error per test over the 16 tests. Errors of substituting meaningful words for nonsense or in reverse were few; 12 Korsakoff *S*s gave records with not a single mistake of this type. Almost all errors were made with monosyllabic words. The experimental *S*s contributed the three exceptions, all dissyllabic meaningful words; two reported nonsense syllables consisting of only two letters, and one gave a three-letter monosyllabic in which two consonants followed each other. The most common error deviating from the model of the test material was a monosyllabic composed of more than three letters.

The recognition tests confirmed the principal findings of the major experiments, i.e., an interpolated and maximally similar task did not reduce—at any rate did not significantly reduce—the gain shown on relearning. The totals of correct recognition were indeed higher after relearning than on first learning for both types of material, though the differences, ranging from .2 to 1.6, were in no instance statistically significant ($p > .05$). In any case, the number of correct choices is no satisfactory test of accurate recognition because an *S* selecting all 30 cards would thus achieve complete success. Therefore, a ratio score was used, the proportion of correct responses to all choices.[2] Means and *SD*s of these scores on the several tests and of gain on relearning are set out in Table 3, which also lists the *t* ratios for means of individual gain scores. As measured by this method, the gains are negative in all instances but one, i.e., relearning made for a smaller ratio of correct choices than recognition after first learning. All four *t* ratios indicate a probability larger than .05, as do those calculated for the mean difference between relearning with and without interference (1.56 for nonsense, .83 for meaningful material). Mean correct recognition averaged over all tests was 5.4 for nonsense and 6.3 for meaningful syllables; the corresponding means of total choices were 8.9 and 10.9. Several *S*s seemed to

TABLE 2

DIFFERENCES BETWEEN RECALL SCORES ON FIRST AND SECOND TEST OF LEARNING

| | Nonsense Syllables | | Meaningful Words | |
|---|---|---|---|---|
| | Mean Differ. | *t* Ratio | Mean Differ. | *t* Ratio |
| Korsakoff group | | | | |
| No interference | .25 | 1.32 | −.15 | 1.00 |
| Retroactive interference | −.10 | .90 | .85 | 2.24 |
| Proactive interference | .28 | .91 | 0 | 0 |
| Control group | | | | |
| No interference | 1.74 | 5.38* | 2.00 | 4.38* |
| Retroactive interference | 1.13 | 3.70* | .56 | 1.53 |
| Proactive interference | 2.00 | 3.78* | 1.67 | 2.29 |

* $p < .01$.

TABLE 3

MEAN RATIOS OF CORRECT TO TOTAL CHOICES IN RECOGNITION TESTS WITH AND WITHOUT INTERFERENCE, AND MEAN GAIN ON RELEARNING

| | Retroactive Interference | | | | No Interference | | |
|---|---|---|---|---|---|---|---|
| | I | II | III | Gain | I | II | Gain |
| Nonsense Syllables | | | | | | | |
| Mean | .72 | .72 | .59 | −.13 | .69 | .70 | .01 |
| *SD* | .18 | .23 | .19 | .28 | .18 | .18 | .18 |
| *t* ratio | | | | 1.89 | | | .29 |
| Meaningful words | | | | | | | |
| Mean | .77 | .74 | .67 | −.10 | .75 | .70 | −.05 |
| *SD* | .14 | .12 | .15 | .23 | .18 | .13 | .16 |
| *t* ratio | | | | 1.93 | | | 1.24 |

[2] The alternative score of correct responses in the first ten would have been appropriate only if *S*s had been advised to restrict themselves to this total; however, such an instruction might have obscured some interference effects, e.g., a more random selection after an interpolated learning task, or the subsequent addition of incorrect responses to some early correct choices. In fact, none of this happened.

know that the learning list contained 10 items and either restricted their choices to this total or remarked that they were unable to recognize that many. Only one exceeded this limit by a wide margin, and did so repeatedly; there were two further instances of a total selection of 14 cards and a few more of 11.

A further test of recognition was inserted in the recall experiment, when *S*s were asked on the second presentation of the learning list whether they had seen it previously and, if the answer was affirmative, when they had seen it. This was done both in the retention and the retroactive interference test. All control *S*s recognized the list as identical with the one seen a few minutes earlier. Only two Korsakoff patients gave the correct answer on either test, while five others said they had seen it before when the question was asked after a period of no intervening activity, and three gave this answer after they had learned a different list. The remaining *S*s either thought they could recognize some of the words or that they had seen the list a week or two earlier.

## DISCUSSION

The experiments confirmed the hypothesis they were designed to test, i.e., while control *S*s show gain on relearning and the effects of retroactive—though not of proactive—interference, Korsakoff patients do not. In view of these findings, an explanation of the amnesic symptoms of Korsakoff psychosis in terms of interference or obliteration by more recent impressions would appear to be not only inadequate but incorrect.

The experiments also furnish some information on the Korsakoff patient's capacities to learn and remember. They confirm the clinical observation that his amnesic defect is neither absolute nor uniform in severity. As tested by correct and more or less immediate recall of words from a list, Korsakoff patients do not form a clearly distinct population; while there is some overlap between their performance and that of the control *S*s, they tend to occupy the lower end of a scale on which persons with no amnesic symptoms cluster higher up. On the other hand, the Korsakoff patients' retentive capacity appears to be nil as tested by gain on relearning, and individual deviations from this norm are negligible. In respect of retention, therefore, they seem to form a discrete population. However, the results of this test do not warrant any wide generalization, and indeed require some qualification in view of an incidental finding. Several Korsakoff patients who showed no gain in relearning nevertheless remembered a few syllables over periods longer than the experimental session. When in Experiment II they were presented with the printed list of meaningful words, they immediately recited "mip," "yof," i.e., two or three of the nonsense syllables they had learned a week or two earlier. Apparently, when their comparatively low upper limit of retention has been reached, a second opportunity to learn does not extend its range; neither does a five-minute learning period in comparison with two minutes. As shown in different types of tasks (Talland & Miller, 1959), Korsakoff patients can match control *S*s in performance as long as the instruction, the information they need to carry, is simple; beyond that range neither practice nor other facilitating conditions help to increase their capacities.

Two other incidental observations are of interest. One is the finding that only a small fraction of the experimental group's reproductions differed from the formal model of the test material, i.e., syllables composed of two consonants with a vowel between them, and that the bulk of the deviant responses still remained monosyllabic. *S*s were not told that the words they would have to learn consisted of three letters or of a single syllable. It would appear that this is something they did learn about the form of the material, even though they failed to learn its content. This finding applies to nonsense and meaningful material alike. Naturally enough, the Korsakoff patients retained—at least for the duration of the test—meaningful words more efficiently than nonsense syllables, but so did the control group. Both groups also made more errors with meaningless material, but this differential effect was significant only in the Korsakoff group. Altogether the Korsakoff patients made more errors, but the significant finding is not their stronger tendency to call an incorrect word or syllable but their relative infrequency of incorrect response, no matter how few items they could recall for certain.

The second incidental observation concerns

the Korsakoff patient's readiness to say, "I cannot remember any more," his realistic moderation both in reproduction and recognition tests, listing well below the correct total of 10 in the one, rarely exceeding it in the other task. This finding is contrary to what would be expected from patients whose clinical description usually lists "confabulation" high among the characteristic symptoms, thereby suggesting an extraordinary readiness to answer questions positively and to create the semblance of knowledge even at the expense of factual accuracy or plausibility. Experience with Korsakoff patients in other experimental as well as in interview situations does not agree with this view but, rather, supports the conclusions of Victor and Adams (1953) that whatever constitutes confabulation belongs to an earlier phase of the disease. It would be incorrect to attribute this absence of confabulatory responses to lack of motivation. In this study, as in others, the Korsakoff patients' motivation ranged between cooperativeness and genuine ego-involvement. The only exceptions were when two patients protested against having to learn meaningless words, but both could be persuaded to go ahead with the task. Many more were the instances of their being worried, perturbed, and annoyed over the evident limitation of their memory. Many more, too, were the incidents of control *Ss* resenting the tedious assignment of rote learning.

Having demonstrated that the subsequent occurrence of events requiring his attention do not necessarily further impair the Korsakoff patient's poor memory of earlier events, it remains to reconcile this conclusion with other observations which indicate that his retention defect is a direct negative function of recency. His retention curve drops very rapidly in the first three minutes and tends to flatten out in the following nine, while that of the control group declines only very slightly and steadily over 12 minutes with the same type of material as used in the experiments here reported. Obviously, time is not an agent that affects memory. Its apparent effect is usually attributed to the interference of events which take place in it. Events differ in their power to cause retroactive interference, and formal or contentual similarity is one of the recognizable criteria of this power. In normal people, a high degree of similarity tends to interfere most with remembering. This is indeed what happened with the control *Ss*, but not with the Korsakoff group. It is reasonable to hypothesize that these patients are subject to different laws of interference, e.g., for them the transition from a particular task to doing nothing or whatever they think of when they are under no specific instruction results in more severe interference than engagement in a similar task, which actually seems to facilitate their orientation toward the original task. Conclusions drawn from other experiments with this group of patients would support such an argument, and its acceptance might extend the applicability of the interference theory to the Korsakoff group. The danger of such reasoning, however, would be to transform the construct of interference from a reasonably objective into a more or less subjective mental event, resting on a circular definition that interference is that which damages retention. At best, such a reinstatement of the interference principle would have to rest on a more general theory of cognition. The results here reported, for instance, would then be reassessed in conjunction with other observations made on the Korsakoff group as an example of a defect in the constructive organization of sequential events.

## SUMMARY

Tests of memorizing nonsense syllables and meaningful words were given to 20 patients with Korsakoff's psychosis and to a control group of 23, composed in part of psychiatric outpatients with a history of alcoholism, in part of hospitalized patients suffering from neurological diseases other than brain damage. Retention and the effects of retroactive and proactive interference were tested by means of reproduction in both groups, and by recognition in the experimental group only. Control *Ss* showed an increment in their recall score after relearning the same list, and this effect was reduced if another list had to be memorized between first and second learning, i.e., there was evidence of retroactive interference. Korsakoff patients showed a more restricted capacity of acquisition, no gain on relearning, and no retroactive interference effects, whether tested by reproduction or recognition. As judged by their errors, they incidentally also

learned the formal characteristics of the test material. Proactive interference was absent or insignificant in both groups.

The experiments show that patients suffering from the amnesic syndrome have a very limited capacity of retention, and even the most favorable learning conditions fail to extend its range. However, the amnesic defect is not complete; some learning does take place, and responses, whether by recognition or recall, are not arbitrary. The findings are presented in evidence against the attribution of the memory defects characteristic of Korsakoff's psychosis simply or primarily to the interference caused by new impressions. At any rate, in this disease in order to account for loss of memory it is necessary to rely on more complex principles of cognition.

## REFERENCES


BÜRGER-PRINZ, H., & KAILA, M. Über die Struktur des amnestischen Symptomenkomplexes. *Ztschr. f.d. ges. Neurol. u. Psychiat.*, 1930, **124,** 553–595.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–267.

GREGOR, A. Beiträge zur Psychopathologie des Gedächtnisses. *Monatschr. Psychiat. Neurol.*, 1909, **25,** 218–255, 339–386.

LIDZ, T. The amnestic syndrome. *Arch. Neurol. Psychiat.*, 1942, **47,** 588–605.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18,** 50–60.

TALLAND, G. A., & MILLER, ALICEMARIE. Perceptual sets in Korsakoff's psychosis. *J. abnorm. soc. Psychol.*, 1959, **58,** 234–240.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

VAN DER HORST, L. Über die Psychologie des Korsakowsyndroms. *Monatschr. Psychiat. Neurol.*, 1932, **83,** 65–83.

VICTOR M., & ADAMS, R. D. The effect of alcohol on the nervous system. *Proc. Assoc. Res. Nerv. Ment. Dis.*, 1953, **32,** 526–573.

WECHSLER, D. A study of retention in Korsakoff psychosis. *Psychiat. Bull. N. Y. State Hosp.*, 1917, **2,** 403–451.

WECHSLER, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

ZANGWILL, O. L. Clinical tests of memory impairment. *Proc. Royal Soc. Med.*, 1942–43, **36,** 576–580.


Received March 12, 1958.

# SOCIAL ISOLATION vs. INTERACTION WITH ADULTS IN RELATION TO AGGRESSION IN PRESCHOOL CHILDREN

WILLARD W. HARTUP AND YAYOI HIMENO[1]

*Iowa Child Welfare Research Station*

SEVERAL recent studies have been concerned with the effects of nonpunitive social isolation and inconsistent adult nurturance on the behavior of young children. Isolation has been studied chiefly by Gewirtz and his associates. These investigators hypothesized that one result of isolating a child from an adult would be to increase the child's behavior to obtain social reinforcers (e.g., verbal approval or attention). Preschool and primary-grade subjects in two studies by Gewirtz and Baer (1958a, 1958b) made more correct marble-dropping responses when given approval after being isolated for 20 minutes than after a period of interacting with the experimenters. Gewirtz (1954) reports elsewhere that children seek attention more frequently when the adult is not immediately available to the child than when the adult is nearby. These results were interpreted as suggesting that *deprivation* of social reinforcers increases the child's behavior "for" these reinforcers.

Hartup (1958) and Rosenblith (1957) employed inconsistent nurturance rather than isolation as an experimental condition. They have suggested that the adult's withdrawal from the child frustrates the child's dependency behavior (in addition to depriving him of social reinforcers) and that the emotion resulting from frustration is related to the changes in behavior observed in the child. They predicted, much as did Gewirtz and Baer, that inconsistent nurturance would be associated with more efficient performance than consistent nurturance on tasks leading to reassurance for the child. In confirmation they found that children mastered a concept-formation problem more readily and performed better on the Porteus mazes after experiencing inconsistent nurturance than after interacting with the experimenter.

It thus appears that disruption of adult-child interaction, whether by isolation or inconsistent nurturance, quite clearly increases a child's dependence. There remains the question as to whether this change is solely a function of deprivation of social reinforcers or whether frustration is also involved. Further study of children's dependence does not offer the best possibility for differentiating deprivation-produced effects from frustration-produced effects of isolation (or inconsistent nurturance) on children's behavior. Presumably *either* deprivation or frustration should increase the frequency of dependency responses or responses "for social reinforcers." The present study is intended to clarify the effects of nonpunitive social isolation on the child by studying children's aggression, a type of child behavior to which a frustration hypothesis is relevant, but to which a deprivation hypothesis is not.

Children's aggressive behavior following a period of social isolation was compared with their aggressive behavior following interaction with an adult. It was assumed (*a*) that frustration is a component of isolation for the child even when the isolation is nonpunitive and (*b*) that frustration is a stimulus commonly eliciting aggression (among other responses) in American children. If these assumptions are tenable, children should be more aggressive following isolation than following interaction. Positive results, of course, would not rule out the occurrence of deprivation effects as a result of nonpunitive isolation; they would suggest only that other motivational changes (frustration) occur.

A second area of interest in the present study has been the relationship between the sex of the interacting adult (especially as it relates to the sex of the child) and the occurrence of aggressive behavior in the child. Sex of experimenter and sex of child have figured in most of the experimental studies on isolation and inconsistent nurturance cited above. Gewirtz and Baer (1958a) and Gewirtz (1954) found that children behaved more frequently "for" the social reinforcers of opposite-sexed experimenters than "for" like-sexed experimenters following isolation or low availability;

[1] Now at Tsuda College, Tokyo, Japan.

this was most marked for boys. Hartup (1958), who used only female experimenters, reported larger differences in performance on learning tasks between consistent and inconsistent nurturance groups of girls than between similar groups of boys. Rosenblith's results (1957) showed boys performing better at maze-copying when the male experimenter *withdrew* attention than when the female withdrew attention or when attention was *given* to boys by either experimenter. No differences were found in this study for girls. Sex of the child in relation to sex of the adult has not, then, produced consistent behavior changes in children subjected to isolation or inconsistent nurturance.

## METHOD

### *Subjects*

The *S*s for this experiment were 24 boys and 24 girls in attendance at the Preschool Laboratories of the Iowa Child Welfare Research Station. At the time of the investigation, the children ranged in age from 4-0 to 5-8; the mean age of the group was 4-10.

*S*s from the preschool population of children over four years of age were randomly assigned to two sequence-of-session groups. The children in each of these sequence groups were then assigned on a random basis to four experimenters, two males and two females.[2] Boys and girls were assigned to subgroups separately so as to place an equal number of both sexes in each group.

Each child participated in two experimental sessions. One involved nonpunitive isolation from the adult experimenter, the other a period of interaction with the experimenter. After each session the child participated in modified doll play. Each child therefore served as his own control in the experiment. Since other studies (Hollenberg & Sperry, 1951; Pintler, Phillips, & Sears, 1946; Sears, 1951) have shown that amount of aggressive behavior typically increases from the first to the second doll play session, sequence of sessions was counter-balanced: the 24 *S*s in Sequence I experienced the isolation condition during their first session, and the interaction condition during their second session. The *S*s in Sequence II were treated in the reverse order: interaction with the experimenter in the first session and isolation from the experimenter in the second session.

### *Experimenters*

The two men and two women *E*s were between 20 and 30 years of age and were either staff or graduate students of the Iowa Child Welfare Research Station. An analysis of variance indicated that individual experimenter effects on the children's aggressive behavior were not significant. Each *S* had the same *E* for both experimental sessions. Sessions were separated by an interval of from five to eight days; one exception of 20 days occurred because of illness.

### *Procedure*

*Isolation Condition.* *E* approached the child in his nursery school group in a warm, friendly manner and asked if the child "would like to see some new toys." When the child's consent was given[3] they left the group; *E* talked to the child during the trip to the experimental room—a period of two to five minutes. When *E* and the child came into the room *E* asked the child to wait for a "few"minutes because he had "something to do first." Then *E* left the room and closed the door. Four toys (two books, a puzzle, a rubber-band game) were within reach if the child wanted to examine them.

*E* remained in a room across a narrow hall and occasionally made noises so that the child could tell that *E* really was "doing something." At the end of 10 minutes, *E* re-entered the room with a collection of doll play materials. The isolation condition used is thus essentially parallel to that used by Gewirtz (1958a, 1958b), except that it lasted for 10 instead of 20 minutes.

Only one child of 55 pretest and experimental *S*s refused to stay alone in the room during this condition, although a number of children indicated concern in the situation (e.g., restless wandering about the room, coming to the door and looking out, nervous oral or manipulatory habits, etc.).

*Interaction Condition.* When *E* and the child arrived in the laboratory room, *E* continued in conversation with the child and suggested playing with the toys. *E* interacted verbally with the child during a 10-minute period. The *E*s were instructed to show interest and involvement in what the child was doing, to smile or laugh when appropriate, to give help and praise, to answer the child's questions, etc. After 10 minutes, *E* told the child that there were more toys. *E* then went to an adjoining room and returned immediately with the doll play materials.

One child of 55 pretest and experimental *S*s wanted to return to preschool during the interaction condition.

*Doll Play.* The play materials used were as follows:

- Plasticine male figure, 7½″ high, ¾″ thick
- Plasticine female figure, 7″ high, ¾″ thick
- Plasticine child figure of indeterminate sex, 4″ high, ½″ thick
- Doll furniture: double bed, circular dining table, kitchen table, kitchen chair, occasional chair
- Child's hammer and four 1″ wire nails
- 8″ wooden mallet
- 6″ wooden smasher
- 6″ wooden roller
- Tongue depressor
- Standard size knife and fork
- Doll nursing bottle

The dolls were stylized figures, firm but malleable.

[2] The authors wish to acknowledge the assistance of Jean Eddy, Frances Horowitz, and Morton Rieber in collecting the data for this investigation.

[3] Three out of 58 children refused to leave the group when approached in this manner.

They were standardized by being molded to a cardboard outline. The materials were displayed on a 20″ x 18″ plywood board. Two spare sets of dolls were kept out of sight until needed in the play session.

At the beginning of the doll play *E* handed the dolls to the child as follows: male, female, child. After the child had identified each doll, *E* said, "You can play with them in any way you like." Timing of the session then began and continued for 10 minutes. During the session, *E*'s interaction with the child followed the pattern of experimenters in other permissive doll play experiments (Caron & Gewirtz, 1951; Hollenberg & Sperry, 1951; Levin & Turgeon, 1957; Pintler et al., 1946; Sears, 1951; Yarrow, 1948). Care was taken that all reassurances to the child were of a general nature; approval given for specific acts was avoided. Whenever a particular doll lost head and limbs or was completely mashed out of shape, the remnants of the figure were removed and a new doll supplied.

A record of each *S*'s aggressive behavior during the doll play was kept by an observer who watched from a booth equipped with one-way screening. Behavior was considered as aggression which involved injury, destruction or chastisement, physical or verbal or both, expressed by the child or through the medium of the dolls (Sears, 1951). Some examples of the children's aggression in this type of doll play are: (*a*) hammering, smashing, cutting, gouging, twisting, tearing apart of the plasticine figures; (*b*) scolding, threatening, (*c*) having a doll become sick or dead, and other of the types of aggression found in standard doll play. The large majority of aggressive behaviors observed in this study were of the type listed in (*a*) above.

Total frequency of aggressive behavior over 10 minutes was recorded for each *S*. The children's aggression varied widely in intensity, but this aspect of behavior was not rated. Single units of aggression were distinguished from one another by (*a*) changes in the object of aggression, (*b*) changes in the agent of aggression, (*c*) changes in materials used to express aggression, (*d*) changes in the type of aggression (e.g., change from striking to gouging, both with the hammer). When one act continued repetitively, an additional unit of aggression was recorded every 15 seconds. Information recorded for each unit of aggression included (*a*) object, (*b*) type, (*c*) time of occurrence.

A single observer was used to record the children's behavior in the 96 sessions. To obtain an estimate of observer reliability a second observer (naive with respect to which experimental condition preceded the doll play) also recorded the sessions of 12 *S*s. In computing reliability it was possible to identify specific units of aggression in the protocols of both observers. The time of occurrence, materials used, and object of the aggression were all noted by the observer. A unit of agreement required like records on these three characteristics from both observers. Per cent agreement between the two observers based on agreements plus disagreements ranged from 68% to 91%. Agreement for the 12 *S*s combined was 82%, based on 466 agreements and disagreements. A less stringent estimate of observer reliability, without regard to the categories mentioned above, but based on total scores, was 93%.

A square-root transformation acted successfully to reduce the skewness and heterogeneity of variance in the distributions of raw frequency scores employed in the major analysis of the data.

## RESULTS AND DISCUSSION

The effects of isolation vs. interaction, order of session, sex of *E* and sex of *S* on the aggressive behavior of the *S*s were studied by means of an extension of Lindquist's Type IV analysis of variance (Lindquist, 1953). The summary of this analysis is presented in Table 1 in which the *F* ratio for isolation vs. interaction is highly significant. The means shown in Table 2 indicate that the frequency of aggressive behavior in the doll play situation was greater after isolation experience than after interac-

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF AGGRESSIVE BEHAVIOR IN DOLL PLAY WITH FOUR FACTORS: ISOLATION EXPERIENCE VS. INTERACTION EXPERIENCE PRECEDING THE DOLL PLAY, SESSION I VS. SESSION II, SEX OF EXPERIMENTER, AND SEX OF CHILD

($N$ = 48)

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Between *S*s | 47 | | |
| Sex of Experimenter | 1 | 2.278 | — |
| Sex of Child | 1 | 26.783 | 4.95** |
| Sex of Experimenter × Sex of Child | 1 | 18.217 | 3.37* |
| Isolation vs. Interaction × Session | 1 | 11.943 | 2.21 |
| Isolation vs. Interaction × Session × Sex of Experimenter | 1 | 4.195 | — |
| Isolation vs. Interaction × Session × Sex of Child | 1 | .171 | — |
| Isolation vs. Interaction × Session × Sex of Experimenter × Sex of Child | 1 | 18.878 | 3.49* |
| Error (b) | 40 | 5.412 | |
| Within *S*s | 48 | | |
| Isolation vs. Interaction | 1 | 10.451 | 10.51*** |
| Session | 1 | .955 | — |
| Isolation vs. Interaction × Sex of Experimenter | 1 | 1.764 | 1.78 |
| Isolation vs. Interaction × Sex of Child | 1 | .139 | — |
| Session × Sex of Experimenter | 1 | .978 | — |
| Session × Sex of Child | 1 | .513 | — |
| Isolation vs. Interaction × Sex of Experimenter × Sex of Child | 1 | .054 | — |
| Session × Sex of Experimenter × Sex of Child | 1 | 3.470 | 3.49* |
| Error (w) | 40 | .994 | |
| Total | 88 | | |

* $p < .10$.
** $p < .05$.
*** $p < .005$.

TABLE 2

MEAN FREQUENCY OF AGGRESSIVE BEHAVIOR ($\sqrt{X}$) FOR BOYS AND GIRLS UNDER MALE AND FEMALE EXPERIMENTERS FOLLOWING ISOLATION EXPERIENCE AND INTERACTION EXPERIENCE IN FIRST AND SECOND DOLL PLAY SESSIONS

($N$ = 48, 24 boys, 24 girls)

| | Isolation | | Interaction[a] | |
|---|---|---|---|---|
| | Session I | Session II | Session I | Session II |
| Boys | | | | |
| Male E | 5.83 | 6.15 | 5.08 | 5.10 |
| Female E | 3.69 | 5.30 | 4.51 | 3.94 |
| Girls | | | | |
| Male E | 2.73 | 5.46 | 3.86 | 2.40 |
| Female E | 4.95 | 3.91 | 3.93 | 3.91 |

[a] The reader should note that treatments was a within-groups effect in this experiment and that the order of sessions was counterbalanced. Thus *Ss* who experienced isolation in their first session experienced interaction in their second session; those who experienced interaction in the first session had isolation in the second session.

tion experience. The data thus bear out the prediction.

It should be noted at this point that a control session involving neither isolation nor interaction was not included in this study.[4] Therefore, it is not known whether the findings resulted from increased aggression due to isolation experience, decreases in aggression resulting from warm interaction with *E*, or a combination of both processes. Children's aggression tends to increase in the presence of permissive adults in experimental doll play, however (Hollenberg & Sperry, 1951; Sears, 1951). It therefore seems unlikely that interaction with *E* prior to doll play reduced aggressiveness in the present experiment. Extrapolation from other doll play experiments is not entirely appropriate, however, since interaction with *E* came before doll play as well as during it in the present case.

The findings show that the consequences of social isolation extend to the behavioral area of children's aggression as well as to dependency; nonpunitive isolation appears to produce complex motivational changes in young children. It may be that increases in aggression are related to frustrating aspects of the isolation situation; there is also the possibility that disruption of the adult-child interaction aroused anxiety or uncertainty in the preschool *Ss*. The exact nature of the motivational changes resulting from isolation are not clear from the present experiment; isolation would appear, however, to affect young children in some way other than by sheer deprivation of social reinforcers.

A statistically significant sex difference in aggressive behavior is shown by the analysis of variance summarized in Table 1. The means in Table 2 show a greater frequency of aggressive behavior in this situation for boys than for girls, a finding that replicates many other doll play studies (e.g., Caron & Gewirtz, 1951; Pintler et al., 1946; Sears, 1951; Yarrow, 1948), and thus affords some validation of the doll play technique used here.

Of borderline significance in the analysis of variance are the following interactions: (*a*) Sex of $E \times$ Sex of $S$, $p < .10$; (*b*) Sex of $E \times$ Sex of $S \times$ Session, $p < .10$; (*c*) Isolation vs. Interaction $\times$ Session $\times$ Sex of $E \times$ Sex of $S$, $p < .10$. Tentative interpretation of these effects may suggest hypotheses for further study.

The means reported in Table 2 show that boys were somewhat more aggressive with males *Es* than with female *Es*. Thus the male adult presumably had greater capacity to incite aggression (or less capacity to inhibit it) than did the female adult. The female adult tended to incite more frequent aggression in girls than did the male adult, but this pattern appears less consistently in the data for girls than in the data for boys.

The borderline triple interaction involving Session, Sex of *E* and Sex of *S* may be interpreted as follows: boys were more aggressive with male *Es* than with female *Es* in both Session I and Session II. Girls, on the other hand, were more aggressive with female *Es* than male *Es*, but only in Session I. In Session II there was no difference in mean frequency of aggression for girls with female as opposed to male experimenters.

This interaction effect appears to vary from treatment condition to treatment condition. The quadruple interaction effect also approaches significance in the analysis of vari-

[4] It may be that a "true" control session or control group cannot be achieved for a study such as this. Neither beginning doll play immediately nor placing child and adult together (but not interacting) are adequate in this regard.

ance. Examination of the means in Table 2 leads to the following interpretation of this interaction: boys were more aggressive with male *E*s than with female *E*s in both Session I and Session II under both isolation and interaction. These differences, however, were largest when Session I involved isolation and Session II involved interaction. Girls were more aggressive with female *E*s than with male *E*s in Session I, this difference being largest when Session I involved isolation. In Session II girls were more aggressive with female *E*s only when the session involved interaction; when Session II involved isolation, girls showed more aggression with male than female *E*s.

There is some indication in these borderline findings that the conditions of the experiment modified the effects of sex membership on the child's aggressiveness. The tendency for children to be more aggressive with like-sex *E*s than opposite-sex *E*s was very apparent only for the 24 children who were isolated in the first session. These *S*s had little opportunity to establish rapport, learn to expect reinforcement for dependency behavior from *E*, etc. When they reached the doll play part of the session it is possible that these *S*s responded to the adult who was present in terms of habits learned outside the experimental situation, perhaps at home; i.e., if the like-sex parent is more threatening and frustrating to many children than the opposite-sex parent at these ages (as is suggested by Freud's description of the Oedipus complex) and the child generalizes responses to frustration from home to laboratory situation, results like those obtained for these *S*s would occur.

The other 24 *S*s began their sessions by interacting with the adult. Perhaps the warm interaction of the first session established expectations of social reinforcement from *E* regardless of *E*'s sex and regardless of expectations for this kind of reinforcement that the child had learned at home. If such were the case, the differential capacity of male and female *E*s to frustrate the child would be reduced. Reduced sex-of-experimenter differences would then be noted in the children's doll play; such was the case for this group of *S*s in the experiment.

The tendency for aggression to increase from first to second doll play session that has been reported in other doll play studies was not observed. Several recent studies (Hollenberg & Sperry, 1951; Levin & Turgeon, 1957; Siegel, 1957) have shown that the particular nature of the experimental conditions can reduce session-to-session changes in amount of aggression; the present results indicate that nonpunitive isolation and interaction before doll play also affect such session changes.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF AGGRESSIVE BEHAVIOR IN DOLL PLAY BASED ON EXPERIENCES OF ISOLATION VS. INTERACTION, TIME INTERVALS, AND SUBJECTS

($N = 48$)

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Isolation vs. Interaction | 1 | 12.48 | 9.27* |
| Time intervals | 3 | 2.74 | 4.44** |
| Subjects | 47 | 7.12 | |
| Isolation vs. Interaction × Time Intervals | 3 | .38 | |
| Isolation vs. Interaction × Subjects | 47 | 1.35 | |
| Time Intervals × Subjects | 141 | .62 | |
| Isolation vs. Interaction × Time Intervals × Subjects | 141 | .68 | |

* $p < .005$; tested against Isolation vs. Interaction × Subjects.
** $p < .005$; tested against Time Intervals × Subjects.

TABLE 4

MEAN FREQUENCY OF AGGRESSIVE BEHAVIOR ($\sqrt{X}$) IN 4 TIME INTERVALS OF TEN-MINUTE DOLL PLAY

($N = 48$)

| Time Interval | Isolation | Interaction |
|---|---|---|
| First 2.5 minutes | 2.03 | 1.70 |
| Second 2.5 minutes | 2.30 | 2.09 |
| Third 2.5 minutes | 2.21 | 1.73 |
| Fourth 2.5 minutes | 2.42 | 2.00 |

Changes over time in the amount of aggression in the two treatment groups was studied by a treatment × *S*s analysis of variance (Lindquist, 1953). Table 3 shows the results of this analysis. The effects of the experimental conditions are significant, duplicating the previous analysis. Table 4 shows that the mean frequency of the *S*s' aggressive behavior is greater in each 2½-minute interval following isolation than in the corresponding interval following interaction. Thus, a single burst of aggression at the beginning of doll play did not restore "emotional equilibrium" to the children in the experiment. On the other hand,

conflict did not delay the occurrence of the isolation-produced increases in aggression. This finding suggests that the motivational changes resulting from nonpunitive isolation have a considerable duration.

Amount of aggression did vary, however, in the different portions of a given session. The effects of time interval are highly significant in the analysis reported in Table 3. Table 4 shows that the greatest change in aggression over time was that between Block 1 and Block 2. Aggression increased between these two blocks. The means show some decrease in frequency of aggression for both groups from Block 2 to Block 3 and a second increase from Block 3 to Block 4. Appropriate $t$ tests showed that the differences between Block 1 and Blocks 2 and 4 were the only ones significant at acceptable levels of significance ($t_{1,2} = 2.85$, $p < .01$; $t_{1,4} = 3.00$, $p < .01$).

## SUMMARY

In an attempt to clarify the effects of social isolation on the motivations of the child, the aggressive behavior of 48 children in a 10-minute modified doll play situation was studied in relation to other experimental conditions of isolation and interaction with adult experimenters, order of conditions, sex of child, and sex of experimenter.

The major findings were: (*a*) the children showed a significantly greater frequency of aggressive behavior in doll play preceded by isolation than in doll play preceded by interaction experience; (*b*) boys were more aggressive than girls; (*c*) there were no differences in frequency of aggression from first to second session; (*d*) frequency of aggression was greater in each 2½-minute interval following isolation than in the corresponding interval following interaction.

The most important implications of these results are (*a*) that frustration effects (as well as deprivation effects) occur in consequence of a child's being isolated in a nonpunitive fashion by an adult; (*b*) sex of child is a significant variable in experimental studies of children's aggression.

## REFERENCES

CARON, A. J., & GEWIRTZ, J. L. An investigation of the effects of sex category of the interacting adult, chronological age and sex of child on aggressive behavior in doll play. *Amer. Psychologist*, 1951, **6**, 307. (Abstract)

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., SEARS, R. R., FORD, C. S., HOVLAND, C. I., & SOLLENBERGER, R. T. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

GEWIRTZ, J. L. Three determinants of attention-seeking in young children. *Child Develpm. Monogr.*, 1954, **19**, No. 2, 1–48.

GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56. (a)

GEWIRTZ, J. L., & BAER, D. M. The effects of deprivation and satiation on behaviors for social reinforcers. *J. abnorm. soc. Psychol.*, 1958, **57**, 165–172. (b)

HARTUP, W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. *Child Develpm.*, 1958, **29**, 191–201.

HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.

LEVIN, H., & TURGEON, VALERIE. The influence of the mother's presence on children's doll play aggression. *J. abnorm. soc. Psychol.*, 1957, **55**, 304–308.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

PINTLER, MARGARET H., PHILLIPS, R., & SEARS, R. R. Sex differences in the projective doll play of preschool children. *J. Psychol.*, 1946, **21**, 73–80.

ROSENBLITH, JUDY F. Attention-withdrawal in relation to the effects of children's learning by imitation. Unpublished doctoral dissertation, Harvard Univer., 1957.

SEARS, PAULINE. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**, No. 6 (Whole No. 323).

SIEGEL, ALBERTA E. Aggressive behavior of young children in the absence of an adult. *Child Develpm.*, 1957, **28**, 371–378.

YARROW, L. J. The effects of antecedent frustration on projective play. *Psychol. Monogr.*, 1948, **62**, No. 6 (Whole No. 293).

Received March 14, 1958.

# PERSONAL MOTIVATIONS AND SELF-CONCEPTIONS ASSOCIATED WITH EFFECTIVENESS AND INEFFECTIVENESS IN EMERGENT GROUPS

PAUL H. MUSSEN AND LYMAN W. PORTER

*University of California, Berkeley*

WHEN a group consists of people who have had little or no previous contact with each other, a person's past accomplishments and status are not likely to play a direct role in the group's perception of him as an effective contributor. However, since members of such groups contribute differentially from the beginning, some will shortly be recognized as leaders, while others will appear to contribute relatively little to the group's progress. Whether or not an individual will be seen as a leader or an effective participant will depend on a number of factors. Among these, his stable and persistent personality characteristics would seem to be especially important.

There is an impressive body of evidence that such characteristics—e.g., self-confidence, sociability, initiative, and surgency—are associated with leader status in *structured* groups (Gibb, 1954; Stogdill, 1948). More recently, the personality correlates of *emergent* leadership have also been studied. For example, Bass and his colleagues have shown that self-confidence, ascendency, and sociability are related to leadership in initially leaderless groups (Bass, McGehee, Hawkins, Young, & Gebel, 1953; Bass, Wurster, Doll, & Clair, 1953). Other research on personality and social status indicates that emergent leaders in authoritarian and equalitarian groups manifest differences in behavior which are presumably related to personality characteristics (Goodrich, 1953; Haythorn, Couch, Haefner, Langham, Carter: 1956a, 1956b). Compared with those who emerge as leaders in authoritarian groups, the leaders of equalitarian groups were "more sensitive to others, more effective leaders, more prone to making suggestions for actions subject to group sanction, and less likely to give direct orders to others" (Haythorn et al., 1957a, p. 210).

The present study was designed to investigate the relationship of underlying motivations and basic self-concepts—as measured by a projective instrument administered before participation in an emergent group—to perceived effectiveness in this situation. Specifically, five hypotheses were tested:

1. Effectiveness in group participation will depend in part on the individual's desire to establish harmonious social relationships with others. It is hypothesized that highly effective individuals will exhibit stronger underlying needs for affiliation than ineffective individuals.

2. An individual's effectiveness will be related to his general orientation toward accomplishment and toward persistent effort and work. It is hypothesized that, compared with ineffective individuals, those who are highly effective will show stronger need Achievement.

3. An individual who is self-assured is more likely to participate actively in a group and to be viewed as an important contributor. Hypothesis 3 states that strong feelings of adequacy will be more characteristic of effective than of ineffective individuals.

4. Individuals with feelings of inferiority, being hesitant about putting forth their own ideas and thoughts for others to observe, will be unable to function satisfactorily in a social situation. It is therefore hypothesized that negative self-concepts will be more characteristic of ineffective than of effective members of the group.

5. If effective performance in a group social situation requires the willingness to share ideas with others and accept others' suggestions, basically aggressive individuals will not be effective in the group. The fifth hypothesis states that relatively ineffective members will have more hostile attitudes than the highly effective members.

## METHOD

### *Subjects*

The experiment involved 10 groups of 6 male undergraduates at the University of California. All 60 *Ss* were enrolled in psychology courses, but had taken less

conflict did not delay the occurrence of the isolation-produced increases in aggression. This finding suggests that the motivational changes resulting from nonpunitive isolation have a considerable duration.

Amount of aggression did vary, however, in the different portions of a given session. The effects of time interval are highly significant in the analysis reported in Table 3. Table 4 shows that the greatest change in aggression over time was that between Block 1 and Block 2. Aggression increased between these two blocks. The means show some decrease in frequency of aggression for both groups from Block 2 to Block 3 and a second increase from Block 3 to Block 4. Appropriate $t$ tests showed that the differences between Block 1 and Blocks 2 and 4 were the only ones significant at acceptable levels of significance ($t_{1,3} = 2.85$, $p < .01$; $t_{1,4} = 3.00$, $p < .01$).

## SUMMARY

In an attempt to clarify the effects of social isolation on the motivations of the child, the aggressive behavior of 48 children in a 10-minute modified doll play situation was studied in relation to other experimental conditions of isolation and interaction with adult experimenters, order of conditions, sex of child, and sex of experimenter.

The major findings were: (*a*) the children showed a significantly greater frequency of aggressive behavior in doll play preceded by isolation than in doll play preceded by interaction experience; (*b*) boys were more aggressive than girls; (*c*) there were no differences in frequency of aggression from first to second session; (*d*) frequency of aggression was greater in each 2½-minute interval following isolation than in the corresponding interval following interaction.

The most important implications of these results are (*a*) that frustration effects (as well as deprivation effects) occur in consequence of a child's being isolated in a nonpunitive fashion by an adult; (*b*) sex of child is a significant variable in experimental studies of children's aggression.

## REFERENCES

CARON, A. J., & GEWIRTZ, J. L. An investigation of the effects of sex category of the interacting adult, chronological age and sex of child on aggressive behavior in doll play. *Amer. Psychologist*, 1951, **6**, 307. (Abstract)

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., SEARS, R. R., FORD, C. S., HOVLAND, C. I., & SOLLENBERGER, R. T. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

GEWIRTZ, J. L. Three determinants of attention-seeking in young children. *Child Develpm. Monogr.*, 1954, **19**, No. 2, 1–48.

GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56. (a)

GEWIRTZ, J. L., & BAER, D. M. The effects of deprivation and satiation on behaviors for social reinforcers. *J. abnorm. soc. Psychol.*, 1958, **57**, 165–172. (b)

HARTUP, W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. *Child Develpm.*, 1958, **29**, 191–201.

HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.

LEVIN, H., & TURGEON, VALERIE. The influence of the mother's presence on children's doll play aggression. *J. abnorm. soc. Psychol.*, 1957, **55**, 304–308.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

PINTLER, MARGARET H., PHILLIPS, R., & SEARS, R. R. Sex differences in the projective doll play of preschool children. *J. Psychol.*, 1946, **21**, 73–80.

ROSENBLITH, JUDY F. Attention-withdrawal in relation to the effects of children's learning by imitation. Unpublished doctoral dissertation, Harvard Univer., 1957.

SEARS, PAULINE. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**, No. 6 (Whole No. 323).

SIEGEL, ALBERTA E. Aggressive behavior of young children in the absence of an adult. *Child Develpm.*, 1957, **28**, 371–378.

YARROW, L. J. The effects of antecedent frustration on projective play. *Psychol. Monogr.*, 1948, **62**, No. 6 (Whole No. 293).

Received March 14, 1958.

# PERSONAL MOTIVATIONS AND SELF-CONCEPTIONS ASSOCIATED WITH EFFECTIVENESS AND INEFFECTIVENESS IN EMERGENT GROUPS

PAUL H. MUSSEN AND LYMAN W. PORTER

*University of California, Berkeley*

WHEN a group consists of people who have had little or no previous contact with each other, a person's past accomplishments and status are not likely to play a direct role in the group's perception of him as an effective contributor. However, since members of such groups contribute differentially from the beginning, some will shortly be recognized as leaders, while others will appear to contribute relatively little to the group's progress. Whether or not an individual will be seen as a leader or an effective participant will depend on a number of factors. Among these, his stable and persistent personality characteristics would seem to be especially important.

There is an impressive body of evidence that such characteristics—e.g., self-confidence, sociability, initiative, and surgency—are associated with leader status in *structured* groups (Gibb, 1954; Stogdill, 1948). More recently, the personality correlates of *emergent* leadership have also been studied. For example, Bass and his colleagues have shown that self-confidence, ascendency, and sociability are related to leadership in initially leaderless groups (Bass, McGehee, Hawkins, Young, & Gebel, 1953; Bass, Wurster, Doll, & Clair, 1953). Other research on personality and social status indicates that emergent leaders in authoritarian and equalitarian groups manifest differences in behavior which are presumably related to personality characteristics (Goodrich, 1953; Haythorn, Couch, Haefner, Langham, Carter: 1956a, 1956b). Compared with those who emerge as leaders in authoritarian groups, the leaders of equalitarian groups were "more sensitive to others, more effective leaders, more prone to making suggestions for actions subject to group sanction, and less likely to give direct orders to others" (Haythorn et al., 1957a, p. 210).

The present study was designed to investigate the relationship of underlying motivations and basic self-concepts—as measured by a projective instrument administered before participation in an emergent group—to perceived effectiveness in this situation. Specifically, five hypotheses were tested:

1. Effectiveness in group participation will depend in part on the individual's desire to establish harmonious social relationships with others. It is hypothesized that highly effective individuals will exhibit stronger underlying needs for affiliation than ineffective individuals.

2. An individual's effectiveness will be related to his general orientation toward accomplishment and toward persistent effort and work. It is hypothesized that, compared with ineffective individuals, those who are highly effective will show stronger need Achievement.

3. An individual who is self-assured is more likely to participate actively in a group and to be viewed as an important contributor. Hypothesis 3 states that strong feelings of adequacy will be more characteristic of effective than of ineffective individuals.

4. Individuals with feelings of inferiority, being hesitant about putting forth their own ideas and thoughts for others to observe, will be unable to function satisfactorily in a social situation. It is therefore hypothesized that negative self-concepts will be more characteristic of ineffective than of effective members of the group.

5. If effective performance in a group social situation requires the willingness to share ideas with others and accept others' suggestions, basically aggressive individuals will not be effective in the group. The fifth hypothesis states that relatively ineffective members will have more hostile attitudes than the highly effective members.

## METHOD

### *Subjects*

The experiment involved 10 groups of 6 male undergraduates at the University of California. All 60 *S*s were enrolled in psychology courses, but had taken less

than two semesters of psychology. They were recruited on a volunteer basis.

*Procedure*

When the six *S*s of a group were seated in the experimental room, *E* passed out a brief personal history form (name, year in school, major, school activities) for each *S* to fill out. Each *S* was asked if he were acquainted with any of the other five. (Almost all *S*s reported no acquaintance with other members of their own group. The few *S*s who had had prior contact with one or more of the other members reported that it was minimal. Therefore, no *S*s were excluded from further participation because of extensive acquaintance with others in their group.)

After filling out the personal data questionnaires, the *S*s were told that they were to participate in a study of "creative thinking," the first part of which would be an individual task, and the second part a group task. They were encouraged to "think spontaneously and express yourselves freely." They were first asked to write stories in response to eight pictures, five Thematic Apperception Test (TAT) cards (4, 6BM, 7, 8, and 12) and three other pictures (young man looking out a doorway, several men seated around a table, and a young man seated at a desk with a book) used by McClelland, Atkinson, Clark, and Lowell (1953) and by Atkinson, Heyns, and Veroff (1954). *S*s had five minutes to write a story in response to each card. They were told to include the following points: (*a*) what's happening, (*b*) what led up to it, (*c*) what the people are thinking and feeling, and (*d*) how the story ends. The scoring of these protocols is discussed under Results.

When *S*s had completed their eight stories, they were introduced to the second part of the study. They were told this would involve creative thinking in a group situation, a group discussion of "teaching introductory psychology by television." This topic was selected because it would be interesting to the *S*s, would give them a chance to express original ideas, and could be discussed without any particular *S* having an advantage because of specialized knowledge about the topic.

*E* stated that "since we are interested in fresh ideas and spontaneous thoughts on this topic, it isn't essential that you know a great deal about psychology as a field. What we are interested in is what you think should be covered in an introductory psychology course and how this could be most effectively taught on television." The importance of the topic was stressed by reference to the changing academic enrollment in the years immediately ahead. The *S*s were told that at the end of the discussion period, each of them would be asked to write his own summary of the group's discussion.

During the course of the discussion *E* kept a record of the frequency of each person's participation. The group discussion was stopped after 20 minutes, and *E* distributed a sociometric questionnaire. He explained that "before you write up the summary statements, we'd like to get some idea of your views about the participation of the other members of the group." *S*s were instructed to rank the other members of the group on each of 10 questions: Who contributed the best ideas to the group? Who seemed to be the most active in participation? Who was best at stimulating discussion? Who was best at explaining his own ideas to others? Who was best at clarifying the ideas of others? Who was the best critic of others' ideas? Who was best at organizing the ideas of the group? Who was most influential in the group's decisions? Who from this group do you think you would like to get to know better as a friend? If you had to choose a leader from this group, whom would you choose?

When *S*s had filled out the 10-item sociometric questionnaires, they wrote five-minute summaries of the group's discussion and then were dismissed.

## RESULTS

For purposes of testing the hypotheses, the most effective and least effective group members were selected from the total population of *S*s. This was done on the basis of two criteria: (*a*) *S* was in the top one-third or bottom one-third of his own group on the basis of his average rank on the 10 sociometric questions; and (*b*) *S*'s average sociometric rank must have been in the top one-third or bottom one-third of all 60 *S*s participating in the experiment. Those who ranked in the top one-third on both counts were considered as "high effective" (HE) *S*s, and those ranked in the bottom one-third on both criteria were "low effective" (LE) *S*s. There were 17 HE *S*s and 16 LE *S*s. It should be stressed that the only criterion of leadership or effectiveness in this study is an individual's influence in this situation as *perceived* by the other members of the group. Since indices of leadership may not be highly intercorrelated, other *S*s might have been in the HE or LE groups if another criterion—e.g., actual contribution to the achievement of the group goal—had been used. The term effectiveness as used here refers only to perceived leadership as measured by sociometric selection.

As noted earlier, the TAT was used to assess the personality structure of the *S*s. Each of the personality variables involved in the hypotheses was translated into a TAT scoring category. Five needs and descriptive categories, each defined as specifically as possible, were used in the analysis of the TAT protocols. The *S*'s score for each TAT category was the total number of stories in which he made a response in that category. Underlying the use of such a scoring scheme is the assumption that the storyteller has identified with the hero:

the hero's needs are the same as his; the description of the hero reflects the storyteller's self-conceptions.

The TAT variables used, together with the summary definitions of them, follow:

1. *n Affiliation* involves three separate kinds of relationships: n Affiliation-a was scored whenever the hero was pictured as a person who enjoyed the company of others or did something to establish or maintain friendly relationships with his peers; n Affiliation-b was scored whenever the hero showed romantic interest or fell in love; and n Affiliation-c was scored when the hero's relationships with his parents were described in positive terms (e.g., they are understanding of, or helpful to, the hero). The total n Affiliation score consisted of the sum of the three n Affiliation subcategories.

2. *n Achievement*: scored whenever the hero strives to attain some highly creditable goal (e.g., to achieve professional status or political office, invent something, create something artistic).

3. *Feelings of adequacy*: scored whenever the hero is described as self-confident, self-assured, satisfied with the way things are going, capable of solving his own problems.

4. *Negative self-concepts*: scored whenever the hero is pictured as a failure, feels disgusted, ashamed, or angry with himself, or is described in unflattering terms (e.g., louse, lunatic).

5. *n Aggression:* scored whenever the hero expresses anger or hostility by fighting, injuring, hating, or arguing with someone; or by intentionally destroying something.

To test the reliability of this scoring, one of the authors (PM) and another psychologist[1] independently scored 15 protocols (120 stories). The percentage of interrater agreement was 87, computed by the usual formula (number of agreements divided by number of agreements plus number of disagreements). In order to eliminate bias, the scoring used in the present study was done "blind," that is, independently of knowledge of *Ss*' sociometric status.

Frequency distributions of the scores of the 33 *Ss* were made for each of the TAT variables. Each distribution was dichotomized as close as possible to the median. Since there were only small variations in scoring in some of the variables (n Affiliation-a, n Affiliation-c, n Achievement), this procedure resulted in unequal division of the group in some cases. For each variable, scores above the dichotomization point were considered high, those below were considered low.

Chi-square analyses were used to test the hypotheses and to ascertain whether, as had been predicted, high scores on n Affiliation, n Achievement, and feelings of adequacy were more characteristic of HE *Ss* than of LE *Ss*, while high scores on negative self-conceptions and n Aggression were more frequent among LE *Ss*. Table 1 summarizes the findings, listing the number of HE *Ss* and LE *Ss* with high scores in each of the variables, the chi-square values obtained, and the levels of significance.

As Table 1 shows, there is general support for the first hypothesis which states that HE *Ss* would be high in outgoing characteristics and affiliative needs. While none of the differences between the HE and LE *Ss* in the three types of n Affiliation are significant, all of them are in the predicted direction, and the two groups differ significantly in the total n Affiliation score. Apparently, those who are perceived as leaders tend to have higher needs to establish and maintain friendly social interactions, but these needs may be expressed in alterna-

TABLE I
NUMBER OF HE *Ss* AND LE *Ss* WITH HIGH SCORES ON TAT VARIABLES

| Variable | Number of High Scorers | | $\chi^2$ | $p$[a] |
|---|---|---|---|---|
| | HE ($N$ = 17) | LE ($N$ = 16) | | |
| n Affiliation-a | 8 | 6 | .31 | ns |
| n Affiliation-b | 10 | 6 | 1.50 | ns |
| n Affiliation-c | 11 | 10 | .02 | ns |
| n Affiliation-total | 13 | 5 | 6.80 | .005 |
| n Achievement | 15 | 7 | 7.34 | .005 |
| Feelings of adequacy | 12 | 5 | 5.11 | .01 |
| Negative self-concepts | 5 | 13 | 8.93 | .001 |
| n Aggression | 8 | 12 | 2.70 | .05 |

[a] In order to test a one-sided hypothesis when chi square has only one degree of freedom, the chi square test must be converted into its equivalent value in terms of a normal unit deviate (Fisher, 1938). The levels of significance reported here were evaluated in these terms.

[1] The authors wish to express their appreciation to Robert Messman for his participation in this study.

tive ways, such as in relationships with peers (n Affiliation-a) or in romantic interests (n Affiliation-b).

More HE *S*s revealed high needs for achievement, as had been predicted by the second hypothesis. Apparently those who emerge as perceived leaders from an initially leaderless group are those with strong personal motivation for achievement.

The data also support the third hypothesis. Strong feelings of adequacy were more characteristic of HE *S*s than of LE *S*s. Compared to individuals lacking in these feelings, those who are self-assured and self-confident are much more likely to participate actively and vigorously in groups of peers whom they did not know previously.

Conversely, according to these data, those who have many negative self-conceptions are seen as taking relatively little part in the group discussion and consequently are regarded by the other group members as relatively ineffectual. This finding confirms Hypothesis 4

More LE *S*s than HE *S*s scored high in n Aggression. It may be inferred, in accordance with the hypothesis, that underlying hostile attitudes tend to inhibit free and active participation in a spontaneous social group situation.

Although the primary focus of this study is on *perceived* effectiveness in a developing group situation, there is impressive evidence that sociometric choices are significantly associated with frequency of participation as recorded by *E*. Fourteen of the 17 HE *S*s were in the upper one-third of their own group in terms of frequency of participation, while 12 of the 16 LE *S*s were in the lower one-third ($\chi^2 = 10.94$, $p = .001$) This result is similar to one of Bass et al. (1953) who found that success in leadership in an initially leaderless group (as judged by observers) was highly correlated with a factor labelled "verbality" or "discussion participation."

## DISCUSSION

Problem-solving groups—even those made up of people who are originally strangers to each other—soon take on a social structure. If, as in the present situation, the participants are initially unacquainted with each other, high status for any given individual cannot depend on his past accomplishments. If the problem to be solved does not require exceptional skills, capacities, or specialized knowledge, status cannot depend primarily on intellectual ability or training. Under such circumstances the individual's status position, as viewed by other members of the group, must be determined by his behavior in that particular group. This, in turn, is a function of the individual's personality structure, particularly his basic motivational structure and self-conceptions.

The studies of Bass and his colleagues (1953) presented evidence that positive feelings about oneself are related to leadership in initially leaderless groups. Those whom observers rated as successful leaders frequently spoke favorably of themselves in "nondirective" interviews conducted after the group discussion session, while nonleaders more often spoke unfavorably of themselves. Successful leadership was also associated with high scores on the Sociability and Ascendance scales of the Guilford-Zimmerman Temperament Survey.

Our findings, based on projective assessments of personality, not only confirm the conclusions of Bass, but also help to clarify the basic personal motivations underlying the leader's behavior and consequently the esteem accorded him. Both our findings and those of Bass seem to suggest that high or low effectiveness in a new group is a specific manifestation of over-all personal adequacy or ego-strength. This is illustrated in the rather clear-cut constellation or pattern of traits—an "ideal type"—associated with high effectiveness in a new group. This consists of self-confidence, relative freedom from deep-seated hostility and aggressive attitudes, and strong personal motivations for achievement and for forming close relationships with others. These are traits which, in our culture at least, are probably reflections of adequacy, flexibility, and good ego-strength. Viewed in this way, the individual's ability to assume an active and effective role in the group may be regarded as one demonstration of his fundamental security and general personal effectiveness. An individual with these characteristics will meet a new social situation—as he does any new problem—actively, vigorously, and confidently.

On the other hand, the person who harbors strong hostility or feels inadequate, inferior, and worthless will tend not to handle new situations and new problems well. Faced with a

problem he has not worked on previously, and a group of people he does not know, he will be relatively inhibited and tend to withdraw. Other participants will observe this, and consider him as relatively ineffectual and a noncontributor to the group. His status, as our sociometric data show, will be low.

Although the genesis and development of these feelings of adequacy or inadequacy are not clear, it may be assumed that, since they are basic aspects of an individual's personality, they have their roots in early successes or failures in social situations. If the individual has been frequently rewarded for outgoingness and active participation in social situations, his responses under such conditions will acquire considerable habit strength and will tend to be generalized to other interactions. In contrast, the individual who has encountered frequent discouragement in social situations will learn to withdraw from such situations, and the withdrawal responses may become generalized. When faced with a new social situation in which his role is not clearly defined, such an individual will avoid active participation.

## SUMMARY

The present study tested five hypotheses concerning the relationship of underlying motivations and self-conceptions to sociometrically measured effectiveness in an emergent group situation. Ten groups of six undergraduates each were used as *S*s. After the subjects individually wrote brief stories in response to eight pictures, they participated in a 20-minute leaderless group discussion.

Following the group discussion, each *S* filled out a sociometric questionnaire about the members of his group. Seventeen highly effective (HE) *S*s and 16 relatively ineffective (LE) *S*s were selected on the basis of an analysis of the sociometric responses. Analyses of the stories of these subjects yielded measures of five needs and self-concept categories. In comparison with the LE *S*s, the HE *S*s showed greater n Affiliation, greater n Achievement, stronger feelings of adequacy, fewer negative self-concepts, and less n Aggression. The results may be interpreted as indicating that effectiveness in an initially unstructured group situation is systematically related to the individual's personal motivations and self-concepts.

## REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

BASS, B. M., MCGEHEE, C. R., HAWKINS, W. C., YOUNG, P. C., & GEBEL, A. Personality variables related to leaderless group discussion behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 120–129.

BASS, B. M., WURSTER, C. R., DOLL, P. A., & CLAIR, D. J. Situational and personality factors in leadership among sorority women. *Psychol. Monogr.*, 1953, **67**, No. 16 (Whole No. 366). Pp. 1–23.

FISHER, R. A. *Statistical methods for research workers.* (7th ed.) Edinburgh: Oliver & Boyd, 1938.

GIBB, C. A. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Pp. 877–920.

GOODRICH, D. E. Some manifestations of aggression in authoritarian and equalitarian discussion groups. Unpublished doctoral dissertation, Univer. of Rochester, 1953.

HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & Carter, L. F. The effects of varying combinations of authoritarian and equalitarian leaders and followers. *J. abnorm. soc. Psychol.*, 1956, **53**, 210–219. (a)

HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & CARTER, L. F. The behavior of authoritarian and equalitarian personalities in small groups. *Hum. Relat.*, 1956, **9**, 57–74. (b)

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* N. Y.: Appleton-Century-Crofts, 1953.

STOGDILL, R. M. Personal factors associated with leadership. A survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.

Received March 20, 1958.

# REGIONAL DIFFERENCES IN ANTI-NEGRO PREJUDICE[1]

THOMAS F. PETTIGREW

*Harvard University*

ABUNDANT evidence has been advanced in support of the widespread belief that white Southerners are typically more prejudiced against Negroes than are whites in the North (Hyman & Sheatsley, 1956; Myrdal, 1944; Prothro, 1952; Samelson, 1945; Sims & Patrick, 1936; Stouffer, Suchman, DeVinney, Star, & Williams, 1949). But does this heightened hostility represent (*a*) more externalizing personality potential for prejudice among Southerners,[2] (*b*) the effects of different cultural norms and pressures, or (*c*) both of these?

Virtually no work has been focused specifically on this question, but the scant, indirect evidence available suggests that sociocultural and social adjustment factors—not externalizing personality factors—may be the crucial correlates of the stronger anti-Negro attitudes of most Southerners. Roper (1946, 1947), for example, twice found the South to be one of the least anti-Semitic regions in the nation, and Prothro (1952) noted that two-fifths of his white adult Louisiana sample harbored intensely unfavorable attitudes toward Negroes together with favorable attitudes toward Jews. Yet if such externalization variables as authoritarianism were the crux of the problem, the South would be highly anti-Semitic as well as anti-Negro, and Southerners who are prejudiced against Negroes would be typically prejudiced against Jews too.

Furthermore, the prejudice-prone personality supposedly arises from a family pattern marked by harsh domineering parents who reveal little understanding of and affection for their children (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950; Harris, Gough, & Martin, 1950). There is no evidence, however, that such a family pattern is more prevalent in the South than in the North (Davis, Gardner, & Gardner, 1941; Dollard, 1937). Thus it is not surprising that the few white Southern samples that have been administered the F scale have yielded means that fall well within the range of means reported for nonsouthern groups by the California investigators (Adorno et al., 1950; Milton, 1952; Smith & Prothro, 1957).[3]

Prothro (1954) has summed it up. "The small amount of evidence already available in the South suggests that factors other than personality type enter into the 'American dilemma.' But here again the need is for more research below the Mason-Dixon line."

The present research is an investigation into anti-Negro prejudice in the southern and northern United States, using identical instruments on comparable groups in the two regions. The study focuses upon the following central hypothesis: *The anti-Negro prejudice complex of the South is more related to sociocultural and social adjustment factors and less related to externalizing personality factors than in the North.* Specific hypotheses derived from this basic hypothesis are presented together with the results that pertain to them.

## METHOD

During the summer of 1955, two experienced interviewers employed standard public opinion polling techniques in eight towns varying in population from

[1] The author wishes to express his appreciation to Gordon W. Allport for his encouragement and advice, to Herbert C. Kelman for his theoretical and editorial suggestions, and to Charles LaMonte for his interviewing assistance. The project was supported by a grant from the Laboratory of Social Relations of Harvard University.

[2] "Externalizing personality potential" refers to one of the three functions of attitudes delineated by Smith, Bruner, and White (1956): externalization, social adjustment, and object appraisal. "Externalization occurs when an individual, often responding unconsciously, senses an analogy between a perceived environmental event and some unresolved inner problem . . . [and] adopts an attitude . . . which is a transformed version of his way of dealing with his inner difficulty." Attitudes can play a social adjustment role by contributing to the individual's identification with, or differentiation from, various reference groups. And attitudes can serve an object appraisal function by aiding in the process of understanding "reality" as it is defined by the culture.

[3] If one stretches his definition of the South to include Oklahoma (and the author does not), then Christie and Garcia's (1951) study becomes the lone exception to this statement.

4,000 to 14,000, four in each region. The white adult respondents were selected by simple random procedures from recently published town directories. Seventy-four per cent of the original sample was located, and of these only 8% refused to be interviewed. The final sample's 186 Southerners were slightly younger, slightly less educated, and more often native born than the 180 Northerners polled—all differences existing between the regions taken as wholes.

The eight communities were not chosen randomly. Since all of them had new directories, they were probably more prosperous and industrial than average. In addition, the southern localities, located in North Carolina and Georgia, were selected to have Negro population ratios ranging from 10—45%. The northern towns, all located in New England, each had less than 1% Negroes.

The interview schedule contained three Likert scales. A 10-item instrument of authoritarianism (F scale), an 8-item scale of anti-Semitism (A-S scale), and a 12-item measure of anti-Negro prejudice (N scale) were presented both orally and on cards interspersed among numerous filler items concerning mass media participation and background questions. The poll purported to be concerned with the effects of the mass media on public opinion, and it seemed largely due to this guise that the blatantly phrased prejudice statements caused no interview break-offs. Background items ascertained not only age, educational level, and occupation, but also political party identification, military service, father's occupation, and church affiliation and attendance.

The F, A-S, and four items of the N scale were drawn from the investigations of the authoritarian personality.[4] These measures produced with the present sample adequate Spearman-Brown corrected split-half reliabilities of +.86, +.82, and +.88 respectively. The A-S and N scales were further divided into "stereotyped-belief" (s-b) and "exclusion-discrimination" (e-d) subscales. A typical stereotyped-belief N statement is, "Most Negroes are happy-go-lucky and irresponsible," and a typical exclusion-discrimination N item is "Most Negroes scarcely deserve the right to vote."

## RESULTS

### *Tests of the Specific Hypotheses*

The first hypothesis was that more anti-Negro prejudice—both stereotyped-belief and exclusion-discrimination aspects—exists in the southern sample. Table 1 shows that this prediction is supported at a high level of significance. The Southerners' opinions contrast most sharply with those of the Northerners on the exclusion-discrimination statements. Item analyses indicate that all 12 N items separate the two regions at the .001 level, but the two most discriminating items deal explicitly with racial segregation.[5]

TABLE 1

REGIONAL DIFFERENCES IN ANTI-NEGRO PREJUDICE AND AUTHORITARIANISM[a]

| Scale | Northern Sample | | Southern Sample | | $t$ | $p$ |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | | |
| Total N scale | 4.49 | 1.54 | 6.64 | 1.93 | +11.81 | <.001 |
| Ns-b subscale | 5.07 | 1.47 | 6.88 | 1.91 | +10.11 | <.001 |
| Ne-d subscale | 3.90 | 1.69 | 6.28 | 1.99 | +12.40 | <.001 |
| F scale | 6.18 | 1.46 | 6.19 | 1.96 | .03 | n.s. |

[a] Scale scores can vary between 0 and 10. They were calculated on the basis of +4 for agree strongly, +3 for agree, +2 for undecided, +1 for disagree, and 0 for disagree strongly for each item, and then the total scale scores were collapsed into 11 categories for machine analysis. Thus, the higher the score, the greater is the amount of prejudice or authoritarianism.

No difference was predicted to occur between the regions in F scale means. As we see in Table 1, virtually no difference does result.[6] Though considerably more anti-Negro, the southern sample is not more authoritarian in orientation.

The third hypothesis predicted that there would be a closer relationship between the Ne-d and Ns-b subscales in the South than in the North. A difference appears in the expected direction, but does not quite attain statistical significance. Ne-d and Ns-b correlate +.617 in the northern sample and +.711 in the southern sample ($p < .06$), but some difference might have been expected in this direction simply because of the larger variances in the southern data (Table 1).

Following directly from the central hypothesis was the prediction that when education and age are held constant, the relationship between authoritarianism and anti-Negro prejudice is closer in the North than in the

[4] The California statements used are: for the F scale, Items 1, 8, 13, 18, 21, 26, 27, 31, 37, and 42 on pages 255–57; for the A-S scale, Items 3, 4, 13, 15, 22, and 24 on pages 68–69, and Items 4 and 15 on page 70; and for the N scale, Items 8, 31, and 34 on page 105, and Item 14 on page 117 (Adorno et al., 1950).

[5] Detailed analyses of six of these N scale items together with the results of specific questions concerning school desegregation asked in this same study have been published previously (Pettigrew, 1957).

[6] With a different scoring system used, these F scale means are not directly comparable to those published by Adorno et al. (1950).

South.[7] Table 2 summarizes the relevant data. The results do not support the hypothesis, but rather indicate that F and N correlate to similar degrees in both regions.[8] The slightly higher southern correlations again probably reflect the restricted variances of the northern sample on both the N and F scales (Table 1).

Another specific hypothesis deals with the generalization of prejudice in the two regions, a phenomenon typical of ethnocentric personalities (Adorno et al., 1950; Allport, 1954; Harding, Kutner, Proshansky, & Chein, 1954; Hartley, 1946). With age and education partialled out again, anti-Semitism and anti-Negro prejudices were expected to correlate more highly in the North than in the South. Table 3 reveals that again there are no differences between the regions. Externalization factors, as measured by both the F and A-S scales, appear to operate in similar ways in the North and South.[9]

The final hypothesis concerns sex differences in prejudice. On the assumption that women, the "carriers of culture," reflect the mores more directly than men, it was predicted that females would be more anti-Negro than men in the South, but not in the North. The results of Table 4 lend support to these expectations. No significant differences appear between the sexes in the North — though females tend to be slightly more prejudiced. In the South, as predicted, women are con-

[7] Education and age must be controlled in such relationships, because they have been shown to be correlates not only of prejudice (Allport, 1954; Harding, Kutner, Proshansky, & Chein, 1954; Hyman & Sheatsley, 1956) but also of authoritarianism (Christie, 1954; MacKinnon & Centers, 1956).

[8] Bass (1955), Christie, Havel, and Seidenberg (1958), and others have shown that item form contributes in part to the correlations between such unidirectional Likert scales as those used in this study. Part of this form variance is probably eliminated from the present relationships with the partialling out of both age and education. In any case, the present concern is the difference between the correlations of the two regions and not their absolute magnitudes.

[9] Correlations between the N and A-S scales are considered as evidence for the externalizing aspect of prejudice because only anti-Negro attitudes are positively sanctioned by southern culture (Roper: 1946, 1947). Thus for a southerner to be intolerant of both Negroes and Jews strongly suggests that these attitudes serve largely a function of externalization—not of social adjustment—for him.

TABLE 2
CORRELATIONS BETWEEN ANTI-NEGRO PREJUDICE AND AUTHORITARIANISM WITH AGE AND EDUCATION HELD CONSTANT[a]

| Variables | Northern Sample | Southern Sample | $z$ | $p$ |
|---|---|---|---|---|
| Total N scale and F scale | +.27 | +.34 | 0.73 | n.s. |
| Ns-b subscale and F scale | +.29 | +.29 | 0.00 | — |
| Ne-d subscale and F scale | +.25 | +.35 | 1.05 | n.s. |

[a] All six of these partial correlations are significantly different from zero correlation at better than the .01 level.

TABLE 3
CORRELATIONS BETWEEN ANTI-SEMITISM AND ANTI-NEGRO PREJUDICE WITH AGE AND EDUCATION HELD CONSTANT[a]

| Variables | Northern Sample | Southern Sample | $z$ | $p$ |
|---|---|---|---|---|
| Total A-S scale and N scale | +.55 | +.58 | 0.42 | n.s. |
| A-Ss-b subscale and N scale | +.49 | +.52 | 0.37 | n.s. |
| A-Se-d subscale and N scale | +.51 | +.47 | 0.50 | n.s. |

[a] All six of these partial correlations are significantly different from zero correlation at better than the .001 level.

TABLE 4
SEX DIFFERENCES IN ANTI-NEGRO PREJUDICE AND AUTHORITARIANISM

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Females | Males | | |
| **Northern Sample** | | | | |
| Total N scale | 4.56 | 4.41 | 0.65 | n.s. |
| Ns-b subscale | 5.09 | 5.06 | 0.15 | n.s. |
| Ne-d subscale | 4.03 | 3.76 | 1.20 | n.s. |
| F scale | 6.28 | 6.08 | 0.92 | n.s. |
| **Southern Sample** | | | | |
| Total N scale | 6.83 | 6.43 | 1.42 | <.08 |
| Ns-b subscale | 7.00 | 6.74 | 0.93 | n.s. |
| Ne-d subscale | 6.55 | 5.96 | 2.03 | <.03 |
| F scale | 6.47 | 5.98 | 2.09 | <.05* |

* Unlike the one-tailed tests for the predicted N scale differences, this F scale test employed a two-tailed test of significance.

siderably more anti-Negro than men, particularly in an exclusion-discriminatory manner. Item analyses indicate that the three N statements significantly separating the sexes in the South are all e-d items. Southern women in the sample oppose the equal voting, employment, and educational rights of Negroes more than southern men in the sample. Not all of these sex differences in the South can be attributed to southern mores, however; Table 4 shows that on the F scale females received significantly higher scores than males in the South.

*Further Analyses of the Data*

In order to test the importance of sociocultural and social adjustment factors in the two regions, matched pair designs controlling for education and age (and usually sex) were employed with a number of social variables.[10] Political party identification, military service, church attendance, and social mobility were all analyzed for their effects on N, F, and A-S scores to learn if these factors are more important in the South than in the North.

Table 5 presents the political identification results for 62 northern pairs and 36 southern pairs matched for education and age. No significant differences are found on the N, F, and A-S scales between northern respondents who identify themselves as independents rather than as Democrats or Republicans. In the South, however, the political independents are considerably more tolerant of Negroes than party identifiers. This difference is not a function of sex, since the 28 same-sex pairs also provide a discrepancy attaining the same level of significance. The southern pairs also tend to differ on the F and A-S scales — independents being somewhat less authoritarian and anti-Jewish.

Male veterans during or since World War II were matched on age and education with male nonveterans and yielded the results provided in Table 6. While veterans in both of the regional samples show a slight tendency to be more authoritarian, the southern veterans are markedly less anti-Negro than other southern males. By contrast, the northern veterans are somewhat more intolerant of Negroes than other northern males.

Respondents who had attended church within the week were matched on education and age with others who had not attended church within the month. Table 7 shows that

[10] All matches were made with a tolerance of one year in formal education completed and of five years in chronological age. When more than one respondent could be matched in a pair, the one used was selected randomly.

TABLE 5
POLITICAL PARTY IDENTIFICATION AND SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Independents | Identifiers | | |
| **Northern Sample[b]** | | | | |
| Total N scale | 4.26 | 4.37 | 0.44 | n.s. |
| F scale | 5.90 | 6.08 | 0.77 | n.s. |
| Total A-S scale | 3.92 | 3.92 | 0.00 | — |
| **Southern Sample[c]** | | | | |
| Total N scale | 5.67 | 7.11 | 4.54 | <.001* |
| F scale | 5.44 | 5.92 | 1.46 | <.20 |
| Total A-S scale | 3.17 | 3.86 | 1.79 | <.10 |

[a] A political identifier is a respondent who considers himself either a Democrat or Republican; an independent specifically refers to himself as such.
[b] 62 pairs matched on education and age (51 same-sex pairs).
[c] 36 pairs matched on education and age (28 same-sex pairs).
* This same level of significance is maintained for the 28 same-sex pairs.

TABLE 6
MILITARY SERVICE AND SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Veterans | Nonveterans | | |
| **Northern Sample[b]** | | | | |
| Total N scale | 4.32 | 3.57 | 1.83 | <.10 |
| F scale | 6.11 | 5.58 | 1.33 | <.20 |
| Total A-S scale | 2.50 | 2.50 | 0.00 | — |
| **Southern Sample[c]** | | | | |
| Total N scale | 4.47 | 6.33 | 3.30 | <.01 |
| F scale | 6.13 | 5.87 | 0.65 | n.s. |
| Total A-S scale | 3.53 | 4.33 | 1.52 | <.20 |

[a] Only veterans during or since World War II were employed in these matches.
[b] 19 male pairs matched on education and age.
[c] 15 male pairs matched on education and age.

TABLE 7
CHURCH ATTENDANCE AND SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Attenders | Nonattenders | | |
| **Northern Sample[b]** | | | | |
| Total N scale | 4.36 | 4.73 | 1.31 | n.s. |
| F scale | 6.31 | 6.09 | 1.09 | n.s. |
| Total A-S scale | 4.20 | 4.25 | 0.21 | n.s. |
| **Southern Sample[c]** | | | | |
| Total N scale | 7.28 | 6.38 | 2.65 | $<.02$[d] |
| F scale | 6.69 | 6.13 | 2.03 | $<.05$[e] |
| Total A-S scale | 5.28 | 4.69 | 1.47 | n.s. |

[a] An attender is defined as one who has been to church within the week; a nonattender has not been within the month.
[b] 55 pairs matched on education and age (50 same-sex pairs).
[c] 39 pairs matched on education and age (31 same-sex pairs).
[d] For the 31 same-sex pairs, $t$ becomes 3.20 ($p < .01$).
[e] For the 31 same-sex pairs, $p$ remains at the same level.

TABLE 8
NORTHERN CHURCH ATTENDANCE BY RELIGION AND SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Attenders | Nonattenders | | |
| **Protestants[b]** | | | | |
| Total N scale | 4.05 | 5.19 | 3.16 | $<.01$[c] |
| F scale | 5.62 | 6.10 | 1.03 | n.s. |
| Total A-S scale | 2.95 | 3.33 | 0.74 | n.s. |
| **Catholics[d]** | | | | |
| Total N scale | 4.38 | 4.00 | 0.74 | n.s. |
| F scale | 6.62 | 6.54 | 0.17 | n.s. |
| Total A-S scale | 3.62 | 2.92 | 0.97 | n.s. |

[a] An attender is defined as one who has been to church within the month; a nonattender has not been within the month.
[b] 21 pairs matched on education and age (12 same-sex pairs).
[c] For the 12 same-sex pairs, $t$ becomes 2.39 ($p < .05$).
[d] 13 pairs matched on education, age, and sex.

TABLE 9
DOWNWARD MOBILITY AND SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Downwards | Stables | | |
| **Northern Sample[b]** | | | | |
| Total N scale | 6.20 | 4.50 | 2.62 | $<.05$ |
| F scale | 6.00 | 6.50 | 1.25 | n.s. |
| Total A-S scale | 4.30 | 2.80 | 1.72 | $<.15$ |
| **Southern Sample[c]** | | | | |
| Total N scale | 3.57 | 6.42 | 6.67 | $<.001$ |
| F scale | 6.14 | 6.14 | 0.00 | — |
| Total A-S scale | 2.17 | 3.67 | 1.77 | $<.20$ |

[a] Downwardly mobile individuals have a manual occupation, but their fathers held a nonmanual occupation; stable workers also hold a manual job, but their fathers held manual jobs too.
[b] 10 pairs matched on education and age (8 same-sex pairs).
[c] 7 pairs matched on education and age (6 same-sex pairs).

no differences occur between these groups in the North, but church attenders in the southern pairs tend to be significantly more anti-Negro and authoritarian than nonattenders. These differences hold up when only the results for same-sex pairs are considered.

Previous research by Chein (Harding et al., 1954) having suggested the necessity of holding separate the major religious groupings to understand the relationship between church attendance and anti-Negro attitudes, separate analyses were made of the Catholic and Protestant groups in the North. A redefinition of church attender was necessary for these tests; attenders within the month were compared with nonattenders within the month. The small number of Jews in both regional samples and of Catholics in the southern sample made further analyses impractical.

Table 8 shows that when age and education are controlled the northern Protestant attenders are considerably more tolerant of the Negro than nonattenders, but they are not significantly less anti-Semitic or authoritarian. No differences appear between the Catholic attending and nonattending pairs.

Following Greenblum and Pearlin (1953), mobility was defined in terms of generational shifts between manual and nonmanual occupations. Thus downward mobility is recorded when a respondent reports that he (or her husband) holds a manual occupation when his father held a nonmanual job. With matches on age and education again, the comparisons between the downwardly mobile and the stable manual workers (whose fathers were also manual workers) can be seen in Table 9. As noted in previous research (Bettelheim & Janowitz, 1950; Greenblum & Pearlin, 1953),

TABLE 10
UPWARD MOBILITY AND SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Upwards | Stables | | |
| **Northern Sample[b]** | | | | |
| Total N scale | 4.20 | 4.07 | 0.31 | n.s. |
| F scale | 6.13 | 6.19 | 0.09 | n.s. |
| Total A-S scale | 3.29 | 3.36 | 0.15 | n.s. |
| **Southern Sample[c]** | | | | |
| Total N scale | 5.75 | 5.18 | 1.59 | <.20 |
| F scale | 5.93 | 5.71 | 0.51 | n.s. |
| Total A-S scale | 3.68 | 3.32 | 0.86 | n.s. |

[a] Upwardly mobile individuals have a nonmanual occupation, but their fathers held a manual occupation; stable nonmanual respondents also hold nonmanual jobs, but their fathers held nonmanual jobs too.
[b] 15 pairs matched on education, age, and sex.
[c] 28 pairs matched on education, age, and sex.

the downwardly mobile in the northern sample tend to be both more anti-Negro and anti-Semitic—though not more authoritarian. With the few southern pairs, however, this trend is reversed. Equal on F scale means, the stable manual workers are more prejudiced—particularly against the Negro—than the downwardly mobile in the southern sample.

The upwardly mobile are compared with stable nonmanual respondents in Table 10. With age and education controlled, the only difference approaching significance involves upward mobility in the South. Southerners in the sample with nonmanual employment whose fathers had held manual employment tend to be slightly more anti-Negro than other nonmanual Southerners.

Matched pair designs were not necessary to test the effects of two additional social variables: education and the Negro-white ratio of a community. Education proves to be a potent factor in southern attitudes toward the Negro. The results in Table 11 indicate that education correlates more highly with the N scales—particularly the stereotyped-belief subscale—among the southern interviewees. That this higher correlation is not largely a function of differential variance in the two regional samples (Table 1) is suggested by the relationships between education and the F and A-S measures. Though these scales also have larger *SD*s in the South than in the North, they do not relate to education in the manner that the N scales do: F and education are associated to a similar degree in the two regions and A-S and education are somewhat more closely associated in the North than in the South.

The final variable tested is the percentage of Negroes living in a southern community. Table 12 shows the discrepancies in scale scores between the two low ratio areas polled in the South (10 and 18% Negroes) and the two high ratio areas (38 and 45% Negroes). Though not different in either authoritarianism or anti-Semitism, the respondents in the high Negro percentage localities are significantly more anti-Negro—particularly in their exclusion-discriminatory attitudes.

TABLE 11
CORRELATIONS BETWEEN EDUCATION, PREJUDICE, AND AUTHORITARIANISM[a]

| Variables | Northern Sample | Southern Sample | $z$ | $p$ |
|---|---|---|---|---|
| Total N scale and education | −.30 | −.47 | 1.90 | <.06 |
| Ns-b subscale and education | −.26 | −.51 | 2.81 | <.01 |
| Ne-d subscale and education | −.31 | −.47 | 1.80 | <.08 |
| F scale and education | −.38 | −.45 | 0.80 | n.s. |
| Total A-S scale and education | −.40 | −.26 | 1.49 | <.15 |

[a] These product-moment correlations were developed by measuring education in terms of years of formal schooling completed. All significance levels are based on two-tailed tests.

TABLE 12
NEGRO RATIO AND SOUTHERN SCALE SCORES[a]

| Scale | Means | | $t$ | $p$ |
|---|---|---|---|---|
| | Low Ratio Southern Towns | High Ratio Southern Towns | | |
| Total N scale | 6.25 | 6.97 | 2.95 | <.01 |
| Ns-b subscale | 6.54 | 7.16 | 2.21 | <.03 |
| Ne-d subscale | 5.86 | 6.63 | 2.78 | <.01 |
| F scale | 6.15 | 6.23 | 0.31 | n.s. |
| Total A-S scale | 3.82 | 3.68 | 0.50 | n.s. |

[a] The two low ratio towns have 10% and 18% Negroes respectively; the two high ratio towns have 38% and 45% Negroes respectively. All significance tests are based on two-tailed tests.

## DISCUSSION

The central hypothesis of the study consists of two parts: (*a*) the increased importance of sociocultural and social adjustment factors and (*b*) the decreased importance of externalization factors in southern compared to northern attitudes toward the Negro. The first of these predictions is supported by the data, the second must be modified. Social variables are more significant in the South, but such an externalization variable as authoritarianism is equally relevant in both regions for the samples polled.

### *Sociocultural and Social Adjustment Factors and Anti-Negro Prejudice*

The anti-Negro norms of the South are reflected on the personality level in social conformity. The maintenance of negative attitudes toward Negroes becomes a mechanism of social adjustment for the white Southerner. And, as the data indicate, conformity is associated with intolerance, deviance with tolerance.

*Conformity.* Four types of southern respondents, whose positions in the social structure should be associated with conformity to the culture's dictates, proved to be more intolerant of Negroes than their counterparts: females, political party identifiers, church attenders, and the upwardly mobile.

Though the females score higher on F, cultural factors are also important in the southern sex discrepancy in prejudice (Table 4). Women are more conforming to the norms and expectations of the culture than men. Thus, in the present data, they are more faithful attenders of church than men, behavior that meets the demands of small-town life. Moreover, the sharpest sex divergence in the South occurs on the e-d subscale of the N—the same subscale that distinguishes best between the total northern and southern samples (Table 1) and between the low and high Negro ratio communities in the South (Table 12). Southern women therefore reflect that aspect of anti-Negro feeling that is more characteristic of the southern region as a whole and of the intensely anti-Negro areas of the South in particular. No significant sex differences appear in the North, where the women conform to a culture that allows greater tolerance of Negroes.

The intolerance of political party identifiers in the South is further evidence of the power of southern mores (Table 5). In a small southern town, where politics is everybody's game, identifying with a major political party is—at least in part—conforming, and steadfastly remaining an independent is deviating from the norm.[11] Conformity in the political realm, then, is related to conformity in the ideological realm. And again the North has no such difference.

The greater prejudice of southern churchgoers serves partly an externalization function—as indicated by their higher F scores—and partly a social adjustment function (Table 7). Going to church in a small town in the "Bible Belt" South is definitely conforming to community norms; and these norms also definitively prescribe attitudes toward Negroes. By contrast, Protestant attendance in the North is related to tolerance (Table 8).

The upwardly mobile form another group of conformers. Forced to be keenly attuned to their culture's dominant ideologies in their successful climb up the social ladder, these Southerners tend to be more prejudiced against Negroes than their nonmanual colleagues (Table 10). No such difference occurs in the northern data.

*Cultural deviance.* It is a common finding of research on prejudice in Western society that uncoventional people have a tendency to be tolerant of out-groups. This relationship is particularly evident in the southern data. Several types of deviants have been discussed as the opposites of conformers—political independents and nonattenders of church. Other potential deviants are the downwardly mobile, veterans, and the highly educated.

As in other northern studies (Greenblum & Pearlin, 1953; Bettelheim & Janowitz, 1950), downward mobility is positively related with bigotry in the northern sample (Table 9). But a striking reversal occurs in the southern sample. Perhaps in a region that emphasizes status and family background, that draws a sharp distinction between "poor whites" and

[11] Being a Republican in these southern communities is not as deviant as one might imagine. Three of the four southern towns polled have favored some Republican candidates in recent elections.

"respectable whites," and that cherishes its aristocratic traditions (Cash, 1941; Davis et al., 1941; Dollard, 1937), the downwardly mobile Southerner learns to reject much of his culture. And rejecting the culture's stress on status and tradition makes it easier to reject also the culture's dictates concerning the Negro.

The tendency of northern veterans to be somewhat more hostile toward Negroes than northern nonveterans is perhaps due to the introduction to southern culture that many of them received while training in one of the numerous military installations in the South (Table 6). On the other hand, southern veterans have experienced in the service their first equal status contact with Negroes and their first exposure to nonsouthern ways of life. And, with the culture's hold on them loosened, these veterans' Negro attitudes tend to change.

Like the southern veterans, the highly educated in the South have had the opportunity to encounter nonsouthern ideas and values. Thus, while education is an important negative correlate of prejudice in the North, it becomes unusually crucial in the South (Table 11).

*Negro ratio.* These consistent social relationships in the southern data have been interpreted in terms of conformity and deviance from the narrowly prescribed mores of small town southern life. A final finding provides further evidence for such an analysis. Not only are Southerners living in low Negro ratio towns significantly more tolerant of the Negro than Southerners living in high ratio towns, but there are no externalization differences between them (Table 12). That is, their F and A-S scores are not different. Moreover, they differ most on the e-d subscale of the N—the same subscale that most clearly distinguishes the two regional samples as wholes (Table 1).

### *Externalizing Personality Factors and Anti-Negro Prejudice*

The differences between the regional samples in attitude toward the Negro cannot be explained in terms of differences in externalizing personality potential for prejudice; their F scale means are quite similar (Table 1). And, contrary to the central hypothesis, externalization factors as indexed by F scores are of equal importance in the anti-Negro attitudes of the northern and southern respondents. The correlations in the two regions between the N scale and either the F scale (Table 2) or the A-S scale (Table 3) do not differ significantly.

These results indicate the existence in the South of a large number of culturally intolerant individuals who subscribe to the typical regional attitudes concerning the Negro but who are not authoritarians. As the South's norms regarding the Negro continue to change, these *latent liberals* can be expected to mirror these changes in their own attitudes.

## SUMMARY AND CONCLUSIONS

Public opinion polling techniques were employed with 366 randomly selected respondents in eight roughly matched communities in the North and South to investigate anti-Negro prejudice. The restricted nature of the small town samples limits their generality, but the tentative results offer a broad hypothesis for more representative future research: *in anti-Negro prejudice, externalizing personality findings are of equal importance in the North and South, and sociocultural and social adjustment factors are considerably more important in the South and account for the sharp differences in the regions.*

The hypothesis has several implications for prejudice theory. Areas like the South with heightened prejudice against a particular out-group should not necessarily be thought of as harboring more authoritarianism; the special conflict may represent the operation of particular historical, cultural, and social factors. Indeed, this type of analysis has already been applied to the Union of South Africa (Pettigrew, 1958).

A second implication concerns the current racial desegregation process. The success of this movement in the South, this hypothesis contends, does not depend on changing the deeply ingrained orientations of authoritarian personalities; rather, it rests on the effectiveness with which racial integration now going on in the South can restructure the mores to which many latent liberal Southerners conform.

A final implication involves the role of psychological analysis in the study of prejudice. While psychological work has concen-

trated almost entirely on the externalization function of prejudice (e.g., the frustration-aggression and the authoritarian personality research), the social adjustment function of prejudice is paramount in the South. This does not mean, as Rose (1956) has argued, that the psychological approach is of little value in explaining the heightened intolerance of the South. Rather, this hypothesis suggests that future advances in the psychology of prejudice will come when the direct personality concomitants of cultural pressures are isolated and better understood.

## REFERENCES


ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W. *The nature of prejudice*. Cambridge, Mass.: Addison-Wesley, 1954.

BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.

BETTELHEIM, B., & JANOWITZ, M. *Dynamics of prejudice*. New York: Harper, 1950.

CASH, W. J. *The mind of the South*. New York: Knopf, 1941.

CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954.

CHRISTIE, R., & GARCIA, J. Subcultural variation in authoritarian personality. *J. abnorm. soc. Psychol.*, 1951, **46**, 457–469.

CHRISTIE, R., HAVEL, J., & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

DAVIS, A., GARDNER, B., & GARDNER, MARY. *Deep South*. Chicago: Univer. Chicago Press, 1941.

DOLLARD, J. *Caste and class in a southern town*. New Haven: Yale Univer. Press, 1937.

GREENBLUM, J., & PEARLIN, L. I. Vertical mobility and prejudice: A socio-psychological analysis. In R. Bendix & S. M. Lipset (Eds.), *Class, status and power*. Glencoe, Ill.: Free Press, 1953.

HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *The handbook of social psychology*. Vol. II. Cambridge, Mass.: Addison-Wesley, 1954.

HARRIS, D. B., GOUGH, H. G., & MARTIN, W. E. Children's ethnic attitudes: II. Relationship to parental beliefs concerning child training. *Child Develpm.*, 1950, **21**, 169–181.

HARTLEY, E. L. *Problems in prejudice*. New York: Kings Crown Press, 1946.

HYMAN, H. H., & SHEATSLEY, P. B. Attitudes toward desegregation. *Scient. Amer.*, 1956, **195**, 35–39.

MACKINNON, W. J., & CENTERS, R. Authoritarianism and urban stratification. *Amer. J. Sociol.*, 1956, **41**, 610–620.

MILTON, O. Presidential choice and performance on a scale of authoritarianism. *Amer. Psychologist*, 1952, **7**, 597–598.

MYRDAL, G. *An American dilemma*. New York: Harper, 1944.

PETTIGREW, T. F. Desegregation and its chances for success: Northern and southern views. *Soc. Forces*, 1957, **35**, 339–344.

PETTIGREW, T. F. Personality and socio-cultural factors in intergroup attitudes: A cross-national comparison. *J. conflict Resolution*, 1958, **2**, 29–42.

PROTHRO, E. T. Ethnocentrism and anti-Negro attitudes in the deep South. *J. abnorm. soc. Psychol.*, 1952, **47**, 105–108.

PROTHRO, E. T. Social psychology of the South: Challenge without response. *J. soc. Issues*, 1954, **10**, 36–43.

ROPER, E. United States anti-Semites. *Fortune*, 1946, **33**, 257–260.

ROPER, E. United States anti-Semites. *Fortune*, 1947, **36**, 5–10.

ROSE, A. M. Intergroup relations vs. prejudice: Pertinent theory for the study of social change. *Soc. Probl.*, 1956, **4**, 173–176.

SAMELSON, BABETTE. The patterning of attitudes and beliefs regarding the American Negro: An analysis of public opinion. Unpublished doctoral dissertation, Radcliffe College, 1945.

SIMS, V. M., & PATRICK, J. R. Attitude towards the Negro of northern and southern college students. *J. soc. Psychol.*, 1936, **7**, 192–204.

SMITH, C. U., & PROTHRO, J. W. Ethnic differences in authoritarian personality. *Soc. Forces*, 1957, **35**, 334–338.

SMITH, M. B., BRUNER, J. S., & WHITE, R. W. *Opinions and personality*. New York: Wiley, 1956.

STOUFFER, S. A., SUCHMAN, E. A., DEVINNEY, L. C., STAR, SHIRLEY A., & WILLIAMS, R. M., JR. *The American soldier: Adjustment during army life*. Vol. I of *Studies in social psychology in World War II*. Princeton: Princeton Univer. Press, 1949.

# RELATIONSHIPS OF MEASURES OF ANXIETY AND EXPERIMENTAL INSTRUCTIONS TO WORD ASSOCIATION TEST PERFORMANCE[1]

IRWIN G. SARASON[2]
*University of Washington*

THE present experiment was concerned with the performance of subjects of differing anxiety levels on a word association task under two instructional conditions. Three measures of anxiety were employed: the true-false Test Anxiety scale (Sarason, 1958), Bendig's (1956) revision of the Taylor Manifest Anxiety Scale (Taylor, 1953), and the Lack of Protection scale (Sarason, 1958), designed to measure separation anxiety and, like the Taylor scale, conceived as a general anxiety index. On the other hand, the Test Anxiety scale taps a specific anxiety, anxiety in test situations. Subjects differing with respect to score on Edwards' Social Desirability (Edwards, 1957; Fordyce, 1956) scale were also compared on the word association task under the two sets of experimental instructions.

The two sets of instructions used in this experiment differed in that one set informed *S*, prior to the word association task, he was about to take a personality test, whereas the other set of instructions was neutral in character. These instructions and the anxiety measures were then related (*a*) to *S*'s ability to remember word associations he had given and (*b*) to the agreement of *S*'s responses with those of a normative group. Predictions were made concerning the relative standing on these response measures of groups differing in anxiety and in the kind of instructions administered. The following considerations and assumptions provided the basis for the predictions.

There is a growing literature which indicates that high as compared with low anxious *S*s differ in the extent to which their performance is disrupted under conditions of threat or stress (Lucas, 1952; Mandler & Sarason, 1952; Raphelson, 1957; Sarason: 1956, 1957a, 1957b, 1957c; Sarason, Mandler, & Craighill, 1952). Available evidence provides some support for what might be called a habit hypothesis or interpretation of anxiety. The term habit refers to individual differences among *S*s in the response tendencies which they bring with them into situations, e.g., the expectation of failure and inadequacy in strange situations. This hypothesis assumes that high anxious *S*s (i.e., high scorers on anxiety questionnaires) differ from low anxious *S*s in that under conditions of personal threat or danger, such as a situation in which the individual feels he is being evaluated, the high anxious individual will emit responses which are detrimental to the successful completion of his task. These interfering anxiety responses may be of two kinds: (*a*) self-verbalizations, e.g., "I can't pass this test," "I am not as good as other people," and (*b*) such autonomic responses as accelerated heart rate and sweating. At present, however, this habit hypothesis cannot be tested directly with respect to *S*'s emission of inappropriate self-verbalizations. The experimenter has no way of knowing what *S* is thinking or saying to himself at any given moment in an experimental situation. It does seem possible, however, following Cronbach and Meehl's (1955) concept of construct validity, to study the construct of interfering responses and their relation to anxiety as measured by tests. The present study was designed with this purpose in mind.

Since *S*'s inability to remember responses given on a word association test may be a reflection, at least in part, of competing response tendencies, it was hypothesized that high anxious individuals emit more interfering responses under conditions where they are told they are about to take a personality test than under neutral instructions. This hypothesis was based on the assumption that one variable measured by anxiety scales is the tendency to emit task-irrelevant responses under conditions

[1] This research was supported in part by grants from the Agnes H. Anderson Research Fund and the Graduate School Research Fund of the University of Washington.

[2] The writer is indebted to Doris A. Dietze, Robert E. Farley, Doris B. Kommers, and Barry M. Shmavonian for their aid in carrying out this study. The subjects used in the study were obtained through the cooperation of S. S. Culbert, R. B. Loucks, and B. B. McKeever.

of personal or ego threat. In comparisons of high, middle, and low anxious *S*s, therefore, the instructional variable was expected to produce relatively little difference in performance between middle and low anxious *S*s, but that the high anxious *S*s under threat would display a significantly lower level of accuracy in reproducing the word association test response words.

It was further assumed that the more specific the anxiety scale is to the situation in which *S* performs, the greater are the predicted differences. Therefore, in reproducing response words under the experimental instructions, the differences between high as compared with low and middle test anxious *S*s should be greater than those between *S*s differing in general anxiety. Greater confidence was invested in the test anxiety measure because the experimental situation was structured as a test situation.

Word association "commonality" scores were devised to study the possibility of differences attributable to anxiety among *S*s in the agreement of their response words with those of a norm group. Do high, middle, and low anxious *S*s differ among themselves in the tendency to respond to 100 stimulus words with response words of high frequency in the norm group? From the viewpoint of social desirability, it seemed likely that high scorers on Edwards' scale would have higher commonality scores than would low scorers. This possibility suggests itself on the assumptions (*a*) that high and low scorers on the Social Desirability scale differ in their tendency to respond on the basis of social desirability and (*b*) that a tendency to give common response words to stimulus words is more socially desirable than is a tendency to give uncommon responses.

## METHOD

*Subjects.* The *S*s were students in introductory psychology courses at the University of Washington. The Word Association Test was administered to groups varying in size from 15 to 40. The composition of the groups was in no way related to the *S*s' scores on the four previously administered anxiety and social desirability scales. Because of this, it was necessary after all the data had been collected to exclude randomly certain *S*s to meet the requirement of proportionality of cell frequencies in the analyses of variance which were performed.

Since all *S*s had not been administered all of the anxiety and social desirability scales, the *N*s for the analyses of variance of the word association data varied depending upon which scale the given analysis was based. Each of the breakdowns on the scores of the four scales was in terms of upper, middle, and lower thirds of the distribution of test scores.

When *S*'s performance was analyzed in terms of the Test Anxiety scale (TA), 144 *S*s were used. There were 27 *S*s in each of the groups at the three anxiety levels under the experimental condition and 21 *S*s in each of the three control or neutral groups (one each for high, middle, and low TA). High TA scores ranged from 7 to 16, middle TA *S*s had scores ranging from 4 to 6, and low TA *S*s had scores ranging from 0 to 3.

The *N* for the analyses of groups differing in scores on the Taylor Manifest Anxiety Scale (MAS) was 276. There were 46 *S*s in each of the six groups studied. The ranges of scores for high, middle, and low MAS were as follows: high, 6–14; middle, 3–5; and low, 0–2.

The Lack of Protection scale (LP) results were based on 102 *S*s with 17 *S*s per cell. High scores ranged from 8–19; middle, 5–7; and low, 0–4.

When *S*'s performance was analyzed in terms of the Social Desirability (SD) scale, 126 *S*s were employed, with 17 *S*s in each of the two high SD groups, 26 *S*s in each of the middle SD groups, and 20 *S*s in each of the low SD groups. The ranges of scores corresponding to high, middle, and low SD were high, 36–39; middle, 30–35; low, 17–29.

*Procedure.* The following instructions, similar to those used by Russell and Jenkins (1954) in their normative study, were administered to all *S*s at the beginning of the word association test. The present study also used the same stimulus words as those used by Russell and Jenkins as well as those authors' commonality norms for students. These instructions were:

> This is one of the studies in verbal behavior being done at Washington. This particular experiment is on free association.
>
> Please write your name on the outside of the paper passed to you.
>
> When you look at these sheets, you will see a list of 100 stimulus words. After each word write the first word that it makes you think of. Start with the first word: look at it: write the word it makes you think of: then go on to the next word.
>
> Use only a single word for each response.
>
> Do not skip any words.
>
> Work rapidly until you have finished all 100 words.

The only difference between the experimental and control groups was that the *S*s in the control groups were given only the instructions above, whereas the experimental groups also received these instructions immediately following administration of the standard instructions:

> As you will see the test you are about to take is a Word Association Test. This is one of the most sensitive personality tests. From your performance on this test we will obtain leads concerning aspects of your personality makeup. In short, we will know much more about what kind of a person you really are after we study your performance on this test.

No attempt was made to time *S*s' responses on the

word association test. In order to assess *Ss*' ability to reproduce their 100 responses on the test, as soon as the first test was completed, they were given a second test with the instruction to give the same response words which they had given on the immediately preceding first administration.

## RESULTS

The results of the experiment were analyzed in terms of three dependent variables. The first measure was the number of times *S*'s responses on the second word association test were different from the responses given on the first administration. This index is the *reproduction discrepancy score.* The second measure was the number of times *S*'s responses on the first test administration agreed with the most frequent response given by Russell and Jenkins's norm group. The third measure was the number of times *S*'s response word on the first administration was the same as a response word which 100 or more *Ss* in the norm group gave to the stimulus. These last two measures are called the *Commonality*$_1$ score and the *Commonality*$_2$ score respectively. The results are presented in terms of the four questionnaire measures: TA, MAS, LP, and SD.

*Test Anxiety.* The 2 × 3 analysis of variance for TA, using the reproduction discrepancy score (the number of disagreements between *Ss*' responses to the two word association test administrations), revealed a significant $F$ at the .05 level due to instructions and an $F$ for the Anxiety × Instructions interaction which approached significance ($.10 > p > .05$). The $F$ for Test Anxiety alone did not approach significance. Table 1 presents the means and standard deviations for the six groups studied. In order to test certain predictions, $t$ tests were performed between the high TA experimental group and the high TA control group and all other *Ss* in the experiment, regardless of experimental condition. It was found that the High TA experimental group had a significantly greater number of disagreements between the two test administrations than both the High TA control group and all other *Ss* combined ($p < .02$ in both cases). When one considers that the middle TA experimental and control groups did not differ significantly and that a similar result obtained for the Low TA experimental and control groups, it appears that the prediction that the High TA experimental group would differ from all other groups has been supported.

The Test Anxiety 2 × 3 analyses of variance for the two commonality scores failed to yield any statistically significant $F$s. For both measures, however, there was a possible tendency ($.10 > p > .05$) for more commonality with the normative group under the control than under the experimental condition.

*Manifest Anxiety Scale.* For the reproduction discrepancy score the only significant $F$ was that for MAS scores ($p < .001$). The High MAS groups displayed a greater degree of disagreement between their responses on the first and second administrations than did the low and middle MAS groups. The latter groups did not differ significantly from each other. There was a tendency towards significance ($.10 > p > .05$) for the Anxiety × Instructions $F$. When the High MAS experimental group was compared with (*a*) all the other groups and (*b*) with the High MAS control group, in both cases the High MAS experimental group showed the greater discrepancy between the two administrations ($p < .01$ and $p < .05$ respectively). Table 2 presents the means and standard deviations for the groups evaluated in the MAS analysis of variance for the reproduction discrepancies.

For the Commonality$_1$ score (*S*'s agreement with the norm group's most "popular" response word) analysis, significant $F$s were obtained for Anxiety (.01 level) and Anxiety × Instructions (.05 level). The middle and low MAS groups showed greater commonality with the norm group than did the high MAS groups. Again the high MAS *Ss* given the experimental instructions showed lower Commonality$_1$ scores than did the high MAS control *Ss* or all groups combined ($p < .02$ in

TABLE 1
MEANS AND STANDARD DEVIATIONS OF REPRODUCTION DISCREPANCY SCORES FOR SIX TEST ANXIETY SCALE GROUPS

| | Experimental Instructions | | Control Instructions | |
|---|---|---|---|---|
| TA | Mean | *SD* | Mean | *SD* |
| High | 19.15 | 11.61 | 11.57 | 6.51 |
| Middle | 13.33 | 6.76 | 14.62 | 10.05 |
| Low | 14.52 | 9.83 | 11.57 | 5.98 |

TABLE 2
MEANS AND STANDARD DEVIATIONS OF REPRODUCTION DISCREPANCY SCORES FOR SIX TAYLOR MANIFEST ANXIETY SCALE GROUPS

| | Experimental Instructions | | Control Instructions | |
|---|---|---|---|---|
| MAS | Mean | *SD* | Mean | *SD* |
| High | 19.80 | 12.45 | 14.85 | 9.80 |
| Middle | 13.83 | 7.60 | 12.30 | 7.50 |
| Low | 11.76 | 9.75 | 13.13 | 8.40 |

TABLE 3
MEANS AND STANDARD DEVIATIONS OF COMMONALITY$_1$ SCORES FOR SIX TAYLOR MANIFEST ANXIETY SCALE GROUPS

| | Experimental Instructions | | Control Instructions | |
|---|---|---|---|---|
| MAS | Mean | *SD* | Mean | *SD* |
| High | 27.02 | 12.30 | 33.72 | 11.57 |
| Middle | 35.43 | 10.67 | 36.30 | 9.16 |
| Low | 34.09 | 11.19 | 32.26 | 12.07 |

TABLE 4
MEANS AND STANDARD DEVIATIONS OF COMMONALITY$_2$ SCORES FOR SIX TAYLOR MANIFEST ANXIETY SCALE GROUPS

| | Experimental Instructions | | Control Instructions | |
|---|---|---|---|---|
| MAS | Mean | *SD* | Mean | *SD* |
| High | 35.80 | 14.51 | 45.72 | 14.88 |
| Middle | 46.74 | 13.26 | 48.67 | 13.23 |
| Low | 44.15 | 15.02 | 43.02 | 13.70 |

TABLE 5
MEANS AND STANDARD DEVIATIONS OF REPRODUCTION DISCREPANCY SCORES FOR SIX LACK OF PROTECTION SCALE GROUPS

| | Experimental Instructions | | Control Instructions | |
|---|---|---|---|---|
| LP | Mean | *SD* | Mean | *SD* |
| High | 19.59 | 12.07 | 12.82 | 5.47 |
| Middle | 11.88 | 6.38 | 11.94 | 5.81 |
| Low | 15.94 | 8.66 | 11.24 | 6.13 |

both cases). Thus, the experimental instructions seemed to affect the high MAS *S*s in a way different from all other *S*s combined. Table 3 contains the means and standard deviations for the six groups in the Commonality$_1$ analysis of variance. Examination of Table 3 reveals that the main effect for Anxiety (lower commonality scores for high than middle and low MAS groups) was largely determined by the High MAS experimental groups.

The MAS analysis in terms of the Commonality$_2$ score (number of times *S* responded with a word which 100 *S*s in the norm group had emitted) revealed significant *F*s for Instructions (.05 level), Anxiety (.005 level), and Anxiety × Instructions (.03 level). The directions of differences for the latter two significant effects were the same as for the Commonality$_1$ analysis. The significant *F* due to Instructions indicated over-all higher commonality scores for the control than for the experimental groups. Table 4 presents the means and standard deviations for the groups compared in the Commonality$_2$ analysis of variance. Examination of Table 4 again indicates inferiority of the high MAS experimental groups to the other groups studied.

*Lack of Protection.* Using the reproduction discrepancy score, the only significant effect was that for Instructions (.025 level) with greater discrepancy scores under the experimental than under the control condition. There was a tendency ($.10 > p > .05$) in the direction of greater discrepancy scores for the high than for the other two anxiety groups. Significantly greater discrepancy scores for the High LP experimental group compared to the High LP control group or to all other groups were revealed by *t* tests ($p < .05$ and $p < .02$ respectively). Table 5 presents the means and standard deviations for the six groups studied in terms of discrepancy scores.

For both commonality scores there were significant main effects for LP (.025 and .05 levels for Commonality$_1$ and Commonality$_2$ respectively). In both cases, the High LP groups had lower commonality scores than did the Middle and Low LP groups.

*Social Desirability.* The 2 × 3 discrepancy Social Desirability analysis revealed only one significant *F*, that for Instructions (.05 level). There was a greater number of discrepancies between the two administrations under the experimental as compared with the control condition. None of the other effects approached significance.

In both commonality score analyses of vari-

ance, there were no statistically significant *F*s, although there were trends in each case in the direction of higher commonality scores under the control than under the experimental condition. In comparing the performance of low, middle, and high SD *S*s, only the low SD experimental *S*s tended to differ from the high and middle SD *S*s by giving lower commonality scores. No such tendency was observed for the low SD controls.

## DISCUSSION

The results of this experiment seem to provide further support for the plausibility of the interfering response or habit interpretation of anxiety. This is particularly interesting since the present study employed a task quite different in nature from the studies previously reported in the area of the effects of anxiety on performance. The discrepancy score results for all three anxiety measures (TA, MAS, and LP) indicated that the high anxious experimental groups had greater discrepancy scores than all of the other *S*s. This trend was as expected, since it had been assumed that high anxious *S*s would be more sensitive to the personal threat implicit in the experimental instructions than would the low and middle anxious *S*s. The result of this heightened reaction to personal threat was conceived as the emission of personalized task-irrelevant responses which interfere with performance on the task.

While the findings are generally supportive of the hypothesis, all of the results were not as predicted. It was anticipated that the Test Anxiety measure would be more sensitive to the personal threat inferable from the experimental instructions than either the Taylor MAS or the Lack of Protection measures. The results are contrary to this anticipation. If any statement can be made, it is that MAS scores appeared to interact more significantly with instructions than did the other anxiety measures. Certainly, Test Anxiety was not a more sensitive measure than other anxiety scales. Although it is *post hoc*, one possibility comes to mind on the basis of the fact that the TA scale is concerned with *S*'s reaction to *intelligence* test situations. The present experimental instructions emphasized the word association test as a *personality* test. Not knowing the extent to which *S*s who are anxious in intelligence tests are also anxious when undergoing personality evaluation, it perhaps was inappropriate to expect the TA scale to be particularly sensitive to *S*'s performance in the present experiment.

Although the high anxious groups appeared to be more sensitive to the experimental instructions, there is a general trend toward greater discrepancy scores and lower commonality scores under the experimental than under the control instructions. It would thus appear possible that the likelihood of emission of interfering responses increases for all *S*s under a condition of implied personal threat. If this is the case, then high anxious individuals may differ from others, not so much in terms of emission of interfering responses per se, but in a greater proneness to interpret situations as threatening ones. In addition, all individuals may respond to perceived threat or stress similarly, but there may be important individual differences with respect to the intensity of the interfering responses emitted.

A noteworthy difference in results for the anxiety measures was the consistently significant main effect for anxiety for all measures for MAS and LP but not for TA. No hypothesis seems to explain this difference. Intercorrelations of the anxiety measures used in this study show *r*'s in the range of .50–.60. Obviously, then, a good portion of the variance of each test is still unaccounted for in each measure of anxiety.

Since it was felt that high commonality scores might conceivably be a reflection of a need to give socially acceptable responses, *S*s low in Social Desirability should have obtained lower commonality scores than middle and high *S*s. This possibility was supported only in the case of low scorers given the experimental personality instructions. The low Social Desirability control *S*s did not differ from the other *S*s in this experiment with respect to commonality scores. Consequently, it is not easy to explain the present finding simply on the basis of a need for social acceptability. In view of very high negative correlations between measures of social desirability and Taylor anxiety (Edwards, 1957), one may guess that whatever variable is tapped by anxiety scales is also, to a considerable extent, tapped by

indices of social desirability. The discussion of construct validity by Cronbach and Meehl seems especially pertinent to a consideration of the meaning of scales like those offered by Taylor (1953) and Edwards (1957). At least two possibilities must be considered. One is that anxiety is the variable common to such tests. The other is that the commonality between these measures lies in their tapping the extent to which an individual has a set to respond in a socially acceptable manner. Which of these two constructions is more useful is an empirical question of considerable importance.

## SUMMARY

The relationship of scores on several personality tests (Test Anxiety, Taylor Manifest Anxiety, Lack of Protection, and Social Desirability) to *S*'s performance on a word association test was studied. Each *S* took the test twice, instructed on the second administration to give the same responses as on the first. Half the *S*s were told at the outset of the test situation that they were about to take a sensitive personality test; the remaining *S*s were not told this and may be regarded as a control group.

It was found that the high anxious groups given the experimental instructions showed lower commonality scores and greater discrepancies in response between the two word association test administrations than the other *S*s in the experiment. Also, the personal threat of the experimental instructions led to higher discrepancy scores and lower commonality scores (as compared to a norm group) regardless of score on the anxiety and social desirability tests. The different measures of anxiety did not always have similar effects in the experimental situation.

The results were discussed in terms of an interfering response interpretation of anxiety.

## REFERENCES

BENDIG, A. W. The development of a short form of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, **20**, 384.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

LUCAS, J. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

RAPHELSON, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–18.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *USN Off. of Naval Res. Tech. Rep.*, 1954, Contract No. N8, Rep. No. 11.

SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SARASON, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166–171. (a)

SARASON, I. G. The effect of anxiety and two kinds of failure on serial learning. *J. Pers.*, 1957, **25**, 383–392. (b)

SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *J. consult. Psychol.*, 1957, **21**, 485–490. (c)

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. C. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Received March 24, 1958.

# ATTRIBUTE CRITERIALITY AND SEX-LINKED CONSERVATISM AS DETERMINANTS OF PSYCHOLOGICAL SIMILARITY[1]

MICHAEL A. WALLACH AND ALBERT J. CARON
*Harvard University*

In a recent paper, Wallach (1958) has suggested that an important determinant of whether a person will recognize a second event as similar to a first is the nature of the contrast class he is implicitly or explicitly considering when presented with that first event.[2] Suppose, for example, that a person is learning to identify Boston Brahmins. He comes across various instances of this class of persons—some are old, others younger; some rich, others not as rich; and so on. In each case he finds out that the person is of old aristocratic Yankee stock, and eventually he learns to make appropriate identifications. But how is he defining this class to himself? Does he accept every attribute of Boston Yankees as equally important or criterial to the class definition, so that he now has learned such people must be Anglo-Saxon in appearance, enunciate a flat "a," dress conservatively, express themselves with terseness and restraint, and so on? Very likely not. It is more probable that a person learns to define this class of persons mainly in terms of those characteristics which have been *absent* from contrasting negative instances—i.e., from people that he had in mind as part of the non-Yankee class.

Consider two middle-class individuals of recent immigrant extraction, one from New York, the other from Boston. Suppose each sees a certain number of Boston Brahmins—the former in New York, the latter in Boston. Both also think about certain other kinds of people in contrast to Boston Brahmins, but these are likely to be different for each. Our New Yorker will think of middle-class, second generation New Yorkers, and our Bostonian, of middle-class, second generation Bostonians: in each case, people like himself. As a result, these two individuals should regard different attributes as crucial for identifying Boston blue bloods: for the New Yorker, a Boston accent and expressive terseness and restraint should be most criterial (in combination, of course, with physical appearance); for the Bostonian, more subtle aspects of manner and dress (e.g., language idiom, drinking habits, etc.) would be most important. In sum, then, the characteristics we take as criterial for defining a class depend not only on the positive exemplars of that class to which we have been exposed, but also on the particular sorts of negative exemplars that have been considered as a contrast. What one takes as a contrast, in turn, may depend on a host of factors—including one's social background.

But what has all this to do with psychological similarity? The paper cited above also suggested that we consider two events similar insofar as we place them in the same class. And this we will do, in turn, only if both exemplify the property or properties that are criterial for defining that class. It follows, then, that if a second event deviates from a first along an attribute regarded as criterial, it is more likely to be judged different from the first than if it deviates along an attribute regarded as noncriterial. When judging similarity, in other words, there should be less tolerance for deviation along an attribute viewed as criterial than along the same attribute when viewed as noncriterial; there should be greater "conceptual conservatism" in the former case than in the latter.

If we assume, then, that attribute criteriality depends on the contrasting negative instances present during learning, and that criteriality in turn affects psychological similarity, we should be able to predict a person's judgments of similarity or dissimilarity of other events to some key event if we know what contrasts were present when he first learned about that key event. Thus, returning to our

[1] This research is part of a project supported by Grant M-1324 from the United States Public Health Service, and under the general direction of Jerome S. Bruner, whom the authors wish to thank for his encouragement and aid. We are also indebted to Rhea Diamond for a critical reading of the manuscript, to Leonard Green for aid in constructing stimulus materials, and to Jean Yeomans for scheduling the *Ss* and for administering paper-and-pencil tests.

[2] The antecedents of this view are treated at length in the paper cited.

New Yorker and Bostonian, suppose each is now exposed to an upper middle-class Bostonian of immigrant origin and of ambiguous appearance, but whose speech carries the immigrant idiom, and is asked if this person is a Boston Brahmin type. If our hypothesis about contrast and similarity is correct, then the New Yorker (sensitized to accent and restraint) will tend to say yes, while the Bostonian (sensitized to speech patterns) will tend to say no. And we expect this even though both persons have seen the same number and variety of Boston blue bloods.

But another aspect of the similarity question yet remains. Suppose we take a large number of middle-class, second generation Bostonians —people who should tend to use speech patterns as criterial for defining Boston Brahmins—and let them see a variety of upper middle-class Bostonians of immigrant origin, ranging from one who speaks in Ivy League idiom to one who retains most of the immigrant idiom. We ask our sample of judges whether each such upper middle-class Bostonian seems like a Boston Brahmin or not. Since we postulate greater conceptual conservatism for criterial than for noncriterial attributes, we expect these persons to refuse to acknowledge similarity between the various upper middle-class Bostonians and Boston Brahmins much more often than middle-class immigrant New Yorkers (who in turn should be maximally "liberal" in acknowledging similarity). But will all the idiom-sensitive Bostonians be equally adamant in denying the upper middle-class Bostonians a resemblance to Brahmins? Or, stated differently, might there not be individual differences in conceptual conservatism?

A strong hint in this regard comes from recent work by Pettigrew[3] on category width. This investigator reports substantial and reliable differences between males and females in the range of attribute-values one is willing to assign to a category when given the *average* attribute-value of the category, males tending to be "broad" categorizers and females to be "narrow" categorizers. The disposition to bind concepts to their most frequent or most familiar instances—to be a narrow categorizer—seems very close to the notion of conceptual conservatism (intolerance of attribute deviation) advanced above. If this is the case, then, when faced with deviation along a criterial dimension, females, i.e., people who tend to categorize narrowly, should also tend to perceive less similarity than males, i.e., people who tend to categorize broadly.

The foregoing discussion hence has led us to two hypotheses concerning the relationship between conceptual conservatism and psychological similarity. The first assumes that tendencies toward conceptual conservatism should be greater for criterial attributes; and the second, that there are persistent individual differences in conceptual conservatism. The two hypotheses may be stated as follows:

1. *Attribute criteriality:* One event is more likely to be judged similar to a standard event if the two differ on some property which, although always present in the standard, has *not* been learned as criterial for defining it, than if this property *has* been learned as criterial for its definition.

2. *Categorizing style:* When presented with events of varying criterial difference from a standard, females should judge fewer of these events as similar to the standard than males, on the ground that the former tend to categorize more narrowly—i.e., to be more conceptually conservative—than the latter.

The present experiment provides a test for each of these predictions.

## METHOD

Seventy-eight public school children in the sixth grade[4] served as *S*s in a procedure based on the paradigm we have just discussed. They were taught about "poggles": a nonsense name for geometric forms with certain characteristics. For all of the children, exactly the same figures were called poggles. For half the *S*s, however, the contrasting figures that were not poggles differed on one attribute, while for the other half they differed on another attribute.

In brief, the experiment involved three parts: (I) an initial concept attainment session in which differential criteriality for poggles was established; (II) a similarity judgment session involving reports of whether figures of varying deviation from poggles were

[3] A paper-and-pencil measure of category width is presented in Pettigrew (1958).

[4] We are greatly indebted to Ernest R. Caverly, Superintendent of Schools, James R. Hobson, Director of Child Placement, and Elgie Clucas, Principal of the Michael Driscoll School in Brookline, Massachusetts, for making available the pupils and facilities of the Michael Driscoll School for this research.

similar to poggles or not; and (III) a perceptual discrimination session entailing an estimation of the number of different figures presented during the similarity judgment session.

### *Establishment of "Angle" and "Cutout" Criteriality Groups*

In the concept attainment session, all *S*s were given the task of learning to predict accurately whether each of a series of figures is or isn't a poggle. The following instructions were provided:

> I'm going to show you some cards, one at a time. Each one has a different figure on it that looks something like this. (*E* showed *S* the bottom one of a stack of 32 cards.) These figures differ in several ways. Now, I have named some of these figures "poggles," and the rest are not poggles. Which are poggles and which not depends on one or more of the differences in the figures. Your job is to figure out how to tell which are poggles and which not—and there's a definite way of telling. After I show you each card, you tell me whether you think it's a poggle or not and I'll tell you whether you're right or wrong. You'll name some incorrectly at the start because you won't have any clues, but after awhile you should be able to name them all correctly.

The array of 32 cards was then presented one at a time in a prearranged but random order, the series being continued through a second run-through if necessary, until *S* had attained the learning criterion of 16 correct predictions of "poggle" or "not poggle" in a row. All *S*s were exposed to the same four noncriterial or noisy attributes. Thus, half the poggles and half the nonpoggles were blue, and half of each were red; likewise, half of each were large figures and half of each were small; half of each were pointing upward and half of each were pointing to *S*'s right; and half of each had a hole punched near the base while the other half did not. For half the *S*s, however (hereafter called the "angle" group or simply Group A), angularity was made criterial for poggles, in that all poggles had a stubby point and nonpoggles a sharp point. For the remaining *S*s, on the other hand (called the "cutout" group or Group C), cutouts were made criterial by providing all poggles with square cutouts at their base corners, while nonpoggles had no cutouts. The criterial attribute for each group, moreover, remained constant for the other group (i.e., in Group A all figures had cutouts, and in Group C all figures had stubby points). In the figures for each group, then, there were five attributes with two values each, and an additional attribute that was constant. The 32 figures used for each group consisted of all possible combinations of attribute-values. Figure 1 presents examples of poggles, of nonpoggles for Group A, and of nonpoggles for Group C.

### *Similarity Judgment and Discrimination Report Sessions*

The second part of the experiment ensued after *S* had learned the poggle concept to criterion. *E* introduced this next task in these words:

> Now I am going to show you some figures one at a time. None of them, some of them, or all of them, may seem like poggles to you. You just learned what a poggle is, and a poggle is still the same thing. For each of the figures, I want you to tell me whether it seems like a poggle to you or not.

In this series 11 new cards were used, presented one at a time. The figures on all of these cards had square

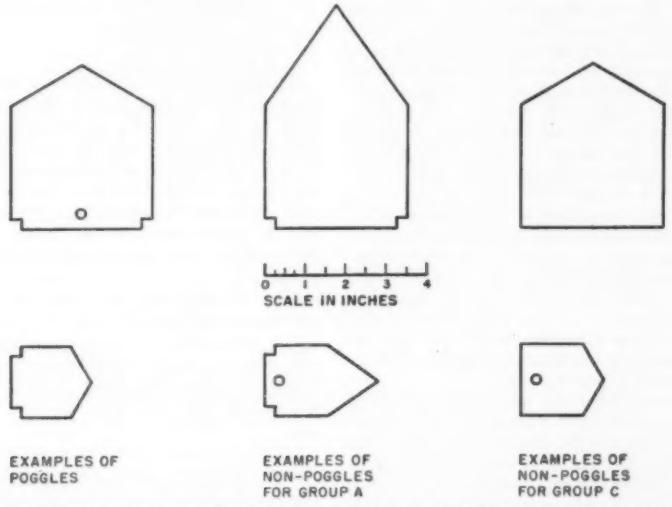


FIG. 1. EXAMPLES OF POGGLES FOR BOTH CONTRAST GROUPS, AND EXAMPLES OF NONPOGGLES FOR EACH OF THE CONTRAST GROUPS, AS USED IN PART I OF THE EXPERIMENT

FIG. 2. THE ELEVEN TEST FIGURES USED IN PART II OF THE EXPERIMENT, SUPERIMPOSED IN ORDER OF INCREASING SHARPNESS OF ANGLE POINT

cutouts from the two corners of their base, were red, large, pointed to *S*'s right, and had no hole punched near the base. The way they varied was in the degree of stubbiness or sharpness of their angle point. The 11 degrees of angle variation used were equal differences in angle starting with one just a bit sharper than the stubby point that had been present in all poggles for both groups, and ending with one just a bit stubbier than the sharp point that had been present in all the contrast instances for Group A. Figure 2 shows these figures superimposed on one another in order of increasing sharpness of angle point. Let us refer to the stubby angle point exemplified by all poggles as Angle 1, and the sharp angle point exemplified by all of Group A's nonpoggles as Angle 13. The angle points used in this second part of the experiment, then, ranged from Angle 2 to Angle 12 inclusive. The first figure presented to *S* was Angle 7—i.e., the angle degree midway between 2 and 12; thereafter the order of the series was identically random for everyone and was gone through twice, for a total of 22 presentations.

To determine whether any differences obtained in similarity judgments might arise from differences in the number of perceptually discriminated angle sizes (Group C perhaps not "seeing" as much angle variation as Group A), *E* asked this final question: "By the way, how many different figures do you think there were in the series I just showed you?"

### *Indices of Conceptual Conservatism*

Our major index for assessing individual differences in conceptual conservatism was the sex of the subject. Nineteen boys and 20 girls constituted Group A, while Group C consisted of 20 boys and 19 girls. Our expectation, on the basis of Pettigrew's findings for adults, that girls would be narrower in category width—or more conceptually conservative—than boys, was confirmed by our finding of a significant sex difference in this direction on the children's form of Pettigrew's "category width" questionnaire,[6] employed as an accessory index in our study.

This written test contained 12 two-part items, and each part of an item was to be answered by choosing from among four alternatives. The questionnaire was called a "Guessing Game," and was introduced with the following instructions, which were both read by the children and repeated orally by the test administrator:

> This game asks you to guess about a lot of things in our world. For instance, if you knew that most grown-up men in the world are around five feet seven inches tall, how tall would you guess the tallest man in the world is? Seven feet tall? Eight feet tall? And how tall would you guess the shortest man in the world is? Three feet tall? Four feet tall? Circle your guesses for each of the things printed below.

The first item was as follows:

> Most birds fly at the speed of about 17 miles per hour.
>
> A. How fast does the fastest bird fly?
>
> 1. 30 miles per hour.
> 2. 21 miles per hour.
> 3. 60 miles per hour.
> 4. 18 miles per hour.
>
> B. How fast does the slowest bird fly?
>
> 1. 15 miles per hour.
> 2. 5 miles per hour.
> 3. 10 miles per hour.
> 4. 2 miles per hour.

Subsequent items concerned the length of whales, the ships in New York harbor, the length of dogs, the speed of cars, the width of roads, the population of states, the height of buildings, the width of windows, the speed of sailboats, schoolbooks written in a year, and amount of daily time spent eating. Each *S*'s category width was measured in the following way: "A" and "B" choices were first coded from 0 to 3, the greater the deviation from the average, the larger the code number; *S*'s score on each item was the sum of his "A" and "B" coded scores for that item, and his total score was the sum of his item scores.

As previously noted, there was a highly significant difference in category width scores between boys and girls in our sample (the medians were 45 and 35, respectively, with $p < .001$ by a two-tail Mann-Whitney test).

---

[6] We are indebted to T. F. Pettigrew for the development and use of a children's form of his category width questionnaire.

### *Dependent Measures and Hypotheses*

Let us consider now the measures provided by our experimental procedure, and the specific predictions implicit in our two hypotheses. The similarity judgment session provides several indices of the range of variation in angle point that a person will tolerate as being similar to a poggle: one measure is the number of trials on which *S* reported the figure was like a poggle (hereafter called frequency or F score), the maximum possible number here being, of course, 22; another measure is the angle of that figure which stands at the median of all the figures which *S* reported as being like a poggle (hereafter called angle or A score); and yet another is whether *S* judged the first figure presented in Part II—namely, the figure whose point was intermediate in angle degree—to be like a poggle (hereafter called the initial figure or I score).

Our attribute criteriality hypothesis leads us to expect that the extent of angle deviation which is still called similar to a poggle, will be greater for Group C than for Group A by each of the three similarity measures just noted. Our hypothesis concerning categorizing style suggests that, for Group A, the extent of angle deviation that is still judged similar to a poggle will be greater for boys than for girls, and be positively correlated with the category width measure, by all measures of similarity. For Group C, on the other hand, where angularity is not criterial, there should be no relation between any differences in similarity range and sex or category width, since the group as a whole should be extreme.

Attempts to reduce any obtained relationship between sex or category width and similarity judgments to some other basis, and hence to imply that the effect is something other than motivational, also had to be evaluated. We had to deal with the possible objection that individual differences in similarity judgments were basically a function of intellectual differences. A measure of general intelligence and a measure of quantitative ability were used in this connection: the Kuhlmann-Anderson IQ (1942), and the average arithmetic achievement score from the Metropolitan Achievement Test (Allen, Bixler, Conner, & Graham, 1946). Both were obtained from our sixth-grade children when they were in the spring term of their fifth-grade year.

## RESULTS

### *Attribute Criteriality and Similarity Judgments*

We find from Table 1 that Groups A and C were not significantly different in their learning of the initial poggle concept (median number of trials to criterion), and also discriminated the same number of different figures during the similarity judgment series (median number of different figures discriminated, or D score). This finding suggests that the attributes used as criterial for defining the concept in each group were of about equal difficulty; and further, that Group A, for whom angle-difference was criterial during concept attainment, was not thereby made more perceptually sensitive to that attribute than Group C. Differences in similarity judgments between the groups, then, cannot be an effect of differing perceptual sensitivities to degree of angle but, rather, must be of conceptual origin.

Further examination of Table 1 indicates such differences in similarity judgments to be considerable. By each of our three measures of similarity (median frequency of similarity calls, or F score; median angle size of the figure standing at the median of all figures *S* reported as being like a poggle, or A score; and number of *S*s judging the initial figure presented in Part II to be like a poggle, or I score), the range of acceptance along the angle dimension is significantly greater for Group C than for Group A. Moreover, the degree of deviation tolerance is not only almost maximal for Group C, but quite low in variability, since we note from Table 1 that the relative frequency of the modal score (or RFM)—a measure that increases as individual differences decrease—is significantly higher in Group C than in Group A for two of the three similarity measures, and in the predicted direction for the third. This smaller variability for

TABLE 1
COMPARISON OF GROUPS A AND C ON VARIOUS EXPERIMENTAL MEASURES

| Measure | Group A | Group C | $p$ |
|---|---|---|---|
| F score | | | |
| median | 9 | 22 | $<.001$* |
| RFM[a] | 7/39 | 27/39 | $<.001$** |
| A score | | | |
| median | 4 | 7 | $<.001$* |
| RFM | 14/39 | 31/39 | $<.001$** |
| I score | | | |
| frequency | 10/39 | 38/39 | $<.001$** |
| RFM | 29/39 | 38/39 | n.s.** |
| Trials to criterion | | | |
| median | 27 | 22 | n.s.* |
| RFM | 9/39 | 5/39 | n.s.** |
| D score | | | |
| median | 3 | 3 | n.s.* |
| RFM | 19/39 | 21/39 | n.s.** |

* Mann-Whitney two-tail test, with corrections for ties, as described by Siegel (1956).

** Chi square two-tail test, with correction for continuity, as described by Siegel (1956).

[a] Relative frequency of the modal score.

Group C than Group A in similarity range indicates that there is but a minimal tendency to consider events nonsimilar when a criterial property is present in both, even though they differ greatly on a noncriterial attribute.

Since these differences occur even though the characteristics present in all positive instances of the poggle class have been the same for both groups, we may conclude that the objective degree of attribute difference between two events does not suffice to determine whether they will appear similar. Rather, psychological similarity depends on whether or not the attribute that varies is criterial for defining either event.

*Categorizing Style and Similarity Judgments*

Turning to our major indicator of categorizing style—the sex of the child—Table 2 shows the sex breakdown for each of the similarity measures. We find that, for Group A, females tolerate less deviation than males by every index. This sex difference in similarity judgments for Group A occurs despite the fact that there is no sex difference in number of trials to attain the poggle concept (and hence no heightened sensitivity to angle due to a difference in number of exposures) nor in the number of figures actually discriminated during the judgment trials. There is no significant sex difference for any measure of similarity in Group C, on the other hand, indicating that what little variance appeared in that condition was due to factors other than conceptual conservatism.

With regard to our accessory index of categorizing style, examination of Table 3 reveals that, for Group A, the correlations between the category width scale and two of our similarity measures (F score and A score) are significant and in the predicted direction, while the correlation of category width with the more conservative I measure, although in the predicted direction, just misses significance. Thus, wide categorizers tend to be more tolerant of deviation in criterial attributes when judging similarity. As with the sex index, there is no significant relationship with category width for any measure of similarity in Group C.

Finally, neither sex nor the category width score is related to IQ or to arithmetic achievement, indicating that the relationships of sex and category width to similarity judgments are not reducible to general intelligence or quantitative factors. (Spearman rank correlations between sex and IQ, sex and arithmetic achievement, category width and IQ, and category width and arithmetic achievement, are all nonsignificant.) Indeed, the Spearman rank correlations between the F measure of similarity (the most representative index) and IQ, and between the F measure and arithmetic achievement, are also nonsignificant for Group

TABLE 2

COMPARISON OF BOYS AND GIRLS WITHIN GROUP A AND WITHIN GROUP C, ON VARIOUS EXPERIMENTAL MEASURES

| Measure | Boys | Girls | $p$ |
|---|---|---|---|
| F score, median | | | |
| Group A | 10 | 7.5 | $< .02$* |
| Group C | 22 | 22 | n.s.* |
| A score, median | | | |
| Group A | 4 | 3.25 | $< .03$* |
| Group C | 7 | 7 | n.s.* |
| I score, frequency | | | |
| Group A | 8/19 | 2/20 | .05** |
| Group C | 20/20 | 18/19 | n.s.** |
| Trials to Criterion, median | | | |
| Group A | 26 | 27 | n.s.* |
| Group C | 21.5 | 26 | n.s.* |
| D score, median | | | |
| Group A | 3 | 3 | n.s.* |
| Group C | 3 | 3 | n.s.* |

* Mann-Whitney two-tail test, with corrections for ties, as described by Siegel (1956).

** Fischer exact probability test, two-tail, as described by Siegel (1956).

TABLE 3

CORRELATIONS* BETWEEN THE CATEGORY WIDTH MEASURE AND SIMILARITY JUDGMENTS IN THE EXPERIMENT

| Measure | Boys and Girls Combined | Boys | Girls |
|---|---|---|---|
| F score | | | |
| Group A | $r = .36$, $p < .03$ | n.s. | n.s. |
| Group C | n.s. | n.s. | n.s. |
| A score | | | |
| Group A | $r = .34$, $p < .04$ | n.s. | n.s. |
| Group C | n.s. | n.s. | n.s. |
| I score | | | |
| Group A | $r = .30$, $p < .07$ | n.s. | n.s. |
| Group C | n.s. | n.s. | n.s. |

* Spearman two-tail rank correlation, with corrections for ties, as described by Siegel (1956).

A and Group C. Further, category width does not reflect any factor critical for similarity other than the sex difference, since, as Table 3 indicates, correlations *within* sex between category width and similarity measures are all nonsignificant.

It thus seems warranted to conclude that individual differences in similarity judgments for Group A were clearly a function of some such factor as conceptual conservatism, while the limited variance in Group C was not a function of this factor.

## DISCUSSION

We have found, then, that psychological similarity depends on attribute criteriality and on individual differences in conceptual conservatism. Let us consider each in turn.

With regard to attribute criteriality, it appears that similarity or dissimilarity of one event to another does not merely depend on the objective degree of difference between them, but rather on whether the one or more properties in whose terms the events differ have been made criterial to *S* for defining either event or not. If criterial, the events are more likely to be judged dissimilar; if not criterial, similar. Criteriality of a property for defining an event, in turn, is established by contrasting that event with others which differ regarding the property in question, so that *S* learns a classification rule to the effect that events exemplifying one value of the property belong in one class and events exemplifying another value in another.

Secondly, we have found that individual differences in conceptual conservatism come into play only when the attribute in question is criterial. What factors determine the limited variance that occurs when the deviating attribute is noncriterial, we cannot say for sure. Considering that for Group C all comparison events contained the attribute-value learned as criterial of poggles (cutout base corners), rejection of similarity would have to involve a shifting of emphasis to a noncriterial property. Whether such rejectors had a clearer perceptual image of the original poggle, or became suspicious of *E*'s intention, or were motivated by a need to be different, we can only conjecture at this point. The one positive statement we can make about this variance is that it had nothing to do with conceptual conservatism.

The relation of sex (and nonrelation of intelligence and quantitative ability measures) both to the range of criterial deviation judged similar and to category width, suggest other factors that may possibly underlie our inferred dimension of conceptual conservatism. There is much evidence in the literature (Komarovsky, 1950; Milner, 1949; Sears, 1951; Sears, Maccoby, & Levin, 1957; Sears, Whiting, Nowlis, & Sears, 1953) indicating that, at least in Western culture, boys in general are more aggressive, more independent, and less docile than girls. While boys tend to be encouraged for independent and aggressive behavior outside the home, girls are usually punished for such behavior. Whereas boys are generally free to choose their own activities and are allowed more privacy in personal affairs, girls tend to be more closely supervised. It seems reasonable to assume that such conditions, implying, as they do, drastic consequences to girls for free expression, render them more restrained and cautious as well as more dependent on external standards.

It is quite possible that such a factor as fear of independent expression may have acted in the present situation to limit the range of similarity acceptance for girls and, also, in Pettigrew's and our own sample, to narrow their category width. In the subculture of experimenter and subject established in our study, a poggle was defined as a form with a particular angle-point (Group A). To allow deviant angle-points into this category is actually to violate the prescriptions of the culture, to take a step beyond convention and authority. Likewise, in Pettigrew's procedure, to assign values to a category beyond those with which one is most familiar, is to rely more on one's own judgment than on accepted standards.

Some corroboration for this position comes from the studies of Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954) and Sandström (1953), which indicated that girls were much more dependent on external context, and much more disturbed by its absence in perceptual situations. The same tendency appears in Sweeney's data (1953) showing that girls are poorer in problem-solving because of a

failure to restructure the field, i.e., to move away from the given.

In sum, we are hypothesizing that for girls there has been a generalization of fear of independence from the sphere of action to that of cognition. Concepts are subject to social regulation as much as behavior, and because girls have been made reluctant to overstep the behavioral boundaries prescribed by authority, so also they are loathe to be expansive in the realm of concepts.

## SUMMARY

Two hypotheses were proposed: 1. *Attribute Criteriality:* One event is more likely to be judged similar to a standard event if the two differ on some property which, although always present in the standard, has *not* been learned as criterial for defining it, than if this property *has* been learned as criterial for its definition. 2. *Categorizing Style:* When presented with events of varying criterial difference from a standard, females should judge fewer of these events as similar to the standard than males, on the ground that the former tend to categorize more narrowly (i.e., conservatively) than the latter.

Attribute criteriality was varied by changing the nature of the negative instances with which positive exemplars of a class were contrasted, and categorizing style was assessed from the sex of the child and also by an independent measure of category width in an experiment involving a concept attainment task and similarity judgments with nonsense figures. Seventy-eight sixth-grade school children served as *S*s, and both hypotheses were sustained. It was concluded that recognition of similarity depends on learned classification rules and on individual differences in conceptual conservatism. The latter in turn seemed due to a sex difference in fear of independent expression.

## REFERENCES

ALLEN, R. D., BIXLER, H. H., CONNER, W. L., & GRAHAM, F. B. *Metropolitan achievement tests: Intermediate battery.* Yonkers, N. Y.: World Book, 1946.

KOMAROVSKY, MIRRA. Functional analysis of sex roles. *Amer. sociol. Rev.*, 1950, **15**, 508–516.

KUHLMANN, F., & ANDERSON, ROSE G. *Kuhlmann-Anderson intelligence tests.* Minneapolis: Educational Publishers, 1942.

MILNER, ESTHER. Effects of sex role and social status on the early adolescent personality. *Genet. Psychol. Monogr.*, 1949, **40**, 231–325.

PETTIGREW, T. F. The measurement and correlates of category width as a cognitive variable. *J. Pers.*, 1958, **26**, 532–544.

SANDSTRÖM, C. I. Sex differences in localization and orientation. *Acta Psychol.*, 1953, **9**, 82–96.

SEARS, PAULINE S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**, No. 6.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

SEARS, R., R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SWEENEY, E. J. Sex differences in problem solving. *Stanford Univer. Depart. of Psychol., Tech. Rep.* No. 1, 1953.

WALLACH, M. A. On psychological similarity. *Psychol. Rev.*, 1958, **65**, 103–116.

WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

Received March 24, 1958.

# RELATIONSHIPS BETWEEN SOCIAL NEED STRIVINGS AND THE DEVELOPMENT OF HETEROSEXUAL AFFILIATIONS[1]

WILLIAM J. MEYER[2]
*Syracuse University*

THE purpose of this investigation was to examine the developmental relationships believed to exist between two social-psychological needs and the social relations structure existing between boys and girls during preadolescence and adolescence. The data to be reported concern boys' and girls' perceptions of the degree to which their same-sex and opposite-sex classmates afford satisfaction of two relatively specific social-psychological needs. Several investigators have reported developmental trends with respect to heterosexual affiliations (Bonney, 1954; Koch, 1944). Generally these studies have shown a decrease in the frequency of opposite-sex choices until grade six and then an increase in such choices. These findings, derived from sociometric tests, suggest that for both sexes the relative ability of the opposite sex to satisfy social needs increases considerably after pubescence. Such an interpretation of these data is somewhat complicated by the fact that the subjects in these studies were required to make choices on the basis of several discrete social situations from which their sociometric scores were computed. This procedure precludes a precise interpretation in terms of the satisfaction of their social needs. Several other important statistical considerations further limit the interpretive value of these studies (see Gardner & Thompson, 1956; Lindzey & Borgatta, 1954).

The present study is based on the assumption that a person is attracted to those group members who are perceived as having a high potential for satisfying one of his social-psychological needs. Rotter (1954) has postulated that a person acquires expectancies "that a particular reinforcement will occur as a function of a specific behavior on his part in a specific situation or situations." The probability that he will seek interactions with another individual in a particular social situation can be ascertained from a knowledge of his previous reinforcements with all the individuals in his group. If the behavior of the group members has been reinforcing, the probability that he will be attracted to them in similar social situations is increased.

Several inferences may be drawn from the foregoing assumptions as they relate to the development of "sex-typed" behavior. Research on this topic indicates that sex appropriate responses to specific social stimuli are well integrated by the sixth year of life (Hattwick, 1937; MacFarlane, Honzik, & Davis, 1937; Tuddenham, 1952). Boys are typically described as outgoing, restless, and mischievous, whereas girls are described as docile, quiet, and conforming. Tuddenham (1952) has shown that preschool children are aware of these sex differences in behavior. Radke (1946) and Sears, Maccoby, and Levin (1957) have shown that parental attitudes towards the limits of acceptable behavior differ for the behavior of sons and daughters and that parents are more demanding and restrictive with their daughters than with their sons. These studies suggest that there are sex differences in the development of psychological needs, an assertion that has received general support in the research literature (Edwards, 1954; Klett, 1957). In addition, these studies suggest that each sex acquires the behavior that is most need-satisfying to its members, and that sex-appropriate social behavior is more clearly defined for girls than for boys.

In this study two relatively specific social needs were investigated for their influence on the development of heterosexual affiliations. The needs of playmirth and succorance were selected because of their relative independence of each other, their relevance to children, and the existence of sex differences in appropriate

[1] This study is an abridged version of a doctoral dissertation submitted to the Psychology Department of Syracuse University. The writer wishes to express his gratitude to George G. Thompson, Chairman of the doctoral committee, and its members, especially Eric F. Gardner and Raymond G. Kuhlen.

[2] Now at the University of Pittsburgh.

social behavior for their reduction.[3] Since the research literature indicates that sex differences exist not only in need strength but also in the behavior required for need reduction, it is anticipated that subjects of each sex will perceive others of the same sex as having a higher potential for satisfying their social need strivings. This sex differential is expected to be maintained throughout the age range included in this study. In addition to the foregoing tentative hypothesis, data relevant to boys' and girls' perceptions of the opposite sex will be examined. One might speculate that since males are allowed more freedom in their social relationships they acquire certain positive expectancies regarding female behavior, which will be reflected in their giving higher opposite-sex sociometric scores.

## METHOD

### *The Sociometric Instrument*

Gardner and Thompson (1956) have published a series of rating scales that overcome some of the psychometric and psychological deficiencies of the traditional sociometric instrument. Underlying these scales is the assumption that the social desirability of a group depends upon the ability of the group members to satisfy specific social-psychological needs. The problem of comparability of social ratings within groups and between groups is resolved by requiring all *Ss* to construct a psychological frame of reference by identifying certain anchor points along a broad psychological continuum. The near equal-interval scaling of this continuum is established from a composite of individual ratings by a factionation procedure.

Each group member is asked to consider a specified hypothetical social situation. He is then required to think of the *one* person out of all the individuals he has ever known, including members of his group, who would be the very best individual to have as a partner for the particular activity. To aid the *Ss* in forming a concept of the *best* individual, a normal distribution curve is provided, with five equal appearing intervals, set off with sufficient space to write in the appropriate names. The *Ss* place the name of the best individual in the box marked *most*. A similar procedure is followed for the *least* liked individual. The preferred-most and preferred-least positions define a psychological continuum common to the members within the group and also common to two or more groups. The next step requires the *Ss* to bisect this continuum into four equal segments by first identifying a *median* individual, and then the individuals who are half-way between the *least* and *median* positions and the *median* and *most* positions. The names, placed at equal intervals along the base line, yield an eight-point scale ranging from slightly above the *least* position to slightly below the *most* position. Ratings of individual group members are made in terms of this reference population by means of a series of forced-choice comparisons.

Separate scales were prepared for each of the two social-psychological needs. To determine the heterosexual social structure for succorance needs, the following hypothetical situation was presented to the *Ss*: "Sometimes you get into trouble and feel unhappy. It might be that you have been blamed for something you didn't do. Think about some time when you were very unhappy and would have liked to talk over your trouble with some kind and sympathetic person." The heterosexual social structure for the playmirth need was assessed by means of the following hypothetical situation: "Suppose you want to have a party. You especially want to invite someone who is always doing things to make people have a good time and have lots of fun." Ten graduate students, making judgments from a large number of such statements, were in unanimous agreement that these statements represent the needs for succorance and playmirth.

The individual ratings of the group members were recorded on a matrix. An average rating *made* by each individual of his classmates was computed by summing across each row of the matrix for each individual and dividing by $N$-1 to obtain the mean. The average rating *received* by each individual from his classmates was computed by summing down each column of the matrix for each individual and dividing by $N$-1.

### *Subjects*

A total of 387 pupils, 212 girls and 175 boys, in Grades 5 through 12 in a rural community in central New York State were administered the scales. The majority of these children were from lower-middle income homes. The distributions of sex, age, and intelligence are not unlike those expected in the average public school. The influence of selection due to school dropouts at the upper grade levels was minimal, since somewhat less than 15% had left school for reasons other than transfer.

### *Procedure*

Test booklets with lists of names appropriate for each classroom were distributed. Four graduate students who were thoroughly familiar with the testing procedures administered the scales. The entire sample was tested on the same day at the end of the sixth month of the school year. The four examiners introduced the test to the pupils as part of a cooperative research study in their school. The average total testing time was approximately 40 minutes. Student cooperation was excellent.

Reliability indices were computed using two sixth-grade classrooms. Since the stability of sociometric scores has been shown to increase with age, it was not considered necessary to determine reliability beyond Grade 6 (Horrocks & Thompson, 1946). Both the need-playmirth and need-succorance scales were readministered two weeks after the initial testing period. Separate indices were determined for ratings given and received.

[3] The definition of these needs follows that of Murray (1938).

The mean reliability for the ratings made on the need-succorance scales is .76; and for the ratings received, .91. For the ratings made and received on need-playmirth, the corresponding values are .69 and .84.

## RESULTS

The hypothesis that a person's perceptions of his own sex are more positive than his perceptions of the opposite sex was tested by comparing the mean ratings made by each sex of their own sex, with the mean ratings made of the opposite sex.[4] Similar analyses were made for both need situations. Since these comparisons involve the same rater making both judgments, the *t* test for matched pairs was used to determine statistical significance (Walker & Lev, 1953). The results of these analyses are summarized in Table 1. As anticipated, the same-sex social ratings are generally higher regardless of the rater's sex. There are no reversals at any of the grade levels for the need-succorance situation though there are two reversals for playmirth. These results lend support to the hypothesis that for each sex, persons of the same sex are best able to reduce social needs. They further suggest that sex differences in ability to satisfy social-need strivings are more clear-cut for succorance than for playmirth behavior.

To test for the presence of a developmental trend in the differences between same-sex and opposite-sex ratings, an analysis of variance of the difference scores was carried out. The null hypothesis concerns the variance between grades. As shown in Table 2, three of the four analyses fail to show a significant effect. The generally nonsignificant between-grades effect can be interpreted as meaning that there is little change in boys' and girls' perceptions of their opposite-sex classmates' ability to satisfy their social needs. For the playmirth situation, however, the data indicate that girls perceive their male classmates as becoming increasingly more capable of satisfying their needs. The sex difference in the developmental trends for the playmirth situation suggests that beginning around Grade 7, girls shift in their perceptions of boys as social companions, though they maintain a greater preference for girls. Boys do not show a similar shift at any grade level.

Another approach to analyzing the data involves a comparison of differences when the sex of the *rater* is different but the sex of the recipient is the same. Since this analysis involves comparisons between independent raters, it was not necessary to analyze difference scores. The analysis of variance is based on the procedure appropriate to categories with unequal *N*s, as described in Walker and Lev (1953). The effect of sex of rater was statistically significant in all four analyses (Table 3). This result further supports the hypothesis that same-sex social interactions are more reinforcing than opposite-sex interactions. The statistically significant differences between grades are a

TABLE 1

MATCHED *t* TESTS BETWEEN MEAN SAME-SEX RATINGS MINUS MEAN OPPOSITE-SEX RATINGS FOR EACH SOCIAL-NEED SITUATION

| Grade placement | Need Succorance | | | | Need Playmirth | | | |
|---|---|---|---|---|---|---|---|---|
| | B×B / B×G | *t* | G×G / G×B | *t* | B×B / B×G | *t* | G×G / G×B | *t* |
| 5 | 3.80 / 3.43 | 2.19* | 4.17 / 2.77 | 4.50** | 3.80 / 3.88 | −.37 | 4.42 / 2.99 | 5.48** |
| 6 | 3.84 / 3.38 | 1.67 | 3.95 / 2.42 | 7.90** | 4.75 / 4.06 | 2.08* | 4.24 / 2.71 | 6.49** |
| 7 | 3.51 / 2.76 | 2.70* | 3.56 / 2.35 | 7.92** | 3.93 / 3.16 | 3.03** | 3.92 / 2.57 | 6.59** |
| 8 | 2.95 / 2.80 | .81 | 3.42 / 2.38 | 4.69** | 3.63 / 3.59 | .17 | 3.76 / 2.84 | 4.74** |
| 9 | 3.18 / 2.36 | 6.01** | 3.48 / 2.24 | 4.88** | 3.74 / 3.06 | 4.15** | 3.47 / 3.28 | .87 |
| 10 | 3.56 / 3.04 | 1.89 | 2.78 / 2.15 | 3.75** | 3.83 / 2.94 | 2.35* | 3.09 / 3.37 | −1.18 |
| 11 | 2.86 / 2.01 | 3.66** | 3.14 / 1.90 | 5.27** | 3.46 / 2.86 | 2.10 | 3.57 / 2.66 | 3.60** |
| 12 | 3.27 / 2.68 | 2.28* | 3.10 / 2.35 | 5.26** | 3.60 / 3.34 | .86 | 3.43 / 2.84 | 2.76* |

* $p < .05$.
** $p < .01$.

[4] The *mean* made and received same-sex and opposite-sex ratings are redundant. To facilitate communication the data reported in this paper are all in terms of ratings made, i.e., a mean rating made by boys of girls.

TABLE 2
ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR EACH SEX COMPARISON AND FOR EACH SOCIAL SITUATION

| Source of Variance | Need Succorance $D_{B\times B-B\times G}$ | | | Need Succorance $D_{G\times G-G\times B}$ | | | Need Playmirth $D_{B\times B-B\times G}$ | | | Need Playmirth $D_{G\times G-G\times B}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *df* | *MS* | *F* | *df* | *MS* | *F* | *df* | *MS* | *F* | *df* | *MS* | *F* |
| Grade | 7 | 1.36 | 1.48 | 7 | 2.51 | 2.34* | 7 | 3.07 | 2.06 | 7 | 11.04 | 7.61** |
| Within | 167 | .92 | | 204 | 1.07 | | 167 | 1.49 | | 204 | 1.45 | |

*$p<.05$.
**$p<.01$.

TABLE 3
ANALYSIS OF VARIANCE OF MEAN SAME-SEX RATINGS AS CONTRASTED WITH MEAN OPPOSITE-SEX RATINGS FOR EACH SOCIAL SITUATION

| Source of Variance | *df* | Need Succorance B × B—G × B | | Need Succorance G × G—B × G | | Need Playmirth B × B—G × B | | Need Playmirth G × G—B × G | |
|---|---|---|---|---|---|---|---|---|---|
| | | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* |
| Grade | 7 | 3.49 | 3.79 | 7.45 | 7.02** | 2.16 | 2.40* | 5.81 | 5.81** |
| Sex | 1 | 102.43 | 111.33** | 36.30 | 34.24** | 77.32 | 85.91** | 13.33 | 13.33** |
| S × G | 7 | .92 | 1.00 | 2.20 | 2.07* | 3.20 | 3.56** | .76 | .76 |
| Within | 370 | .92 | | 1.06 | | .90 | | 1.00 | |

*$p<.05$.
**$p<.01$.

somewhat more complex issue and will be discussed below. The interaction terms for the need succorance (B × B—G × B) and need-playmirth (G × G—B × G) social situations are statistically significant. Inspection of the appropriate means for the playmirth analysis (see Table 1) indicates that the interaction can be attributed to a positive change in the opposite-sex social ratings of the girls after the seventh grade. After Grade 6, girls perceive the boys as increasing in their ability to satisfy playmirth needs, whereas the boys perceive their same-sex classmates as maintaining their same relative ability. A similar trend is not apparent for the opposite-sex ratings of the boys. The significant interaction term for the need-succorance situation (G × G—B × G) appears to be a function of both a decrease in same-sex ratings and an increase in cross-sex ratings.

A more sensitive test of this latter generalization is achieved by an analysis of the opposite-sex ratings (B × G—G × B) for each need situation, as shown in Table 4. The significant interaction for playmirth can be interpreted as meaning that the differences between boys' and girls' perceptions of the opposite sex decrease with increasing grade placement. Figure 1 shows that the decrease in the difference scores can be attributed to the increasingly higher ratings of the girls after Grade 7 in contrast to the relatively constant opposite-sex rating made by the boys. The nonsignificant interaction term for the succorance situation can be interpreted as meaning that the same relative distance between the sexes is maintained over the eight grades studied.

TABLE 4
ANALYSIS OF VARIANCE OF MEAN OPPOSITE-SEX RATINGS MADE BY EACH SEX ON EACH SOCIAL-NEED SITUATION

| Source of Variance | *df* | Need Playmirth B × G − G × B | | Need Succorance B × G − G × B | |
|---|---|---|---|---|---|
| | | *MS* | *F* | *MS* | *F* |
| Grade | 7 | 2.32 | 2.67* | 5.36 | 6.30** |
| Sex | 1 | 18.44 | 21.19** | 21.11 | 24.83** |
| S × G | 7 | 3.90 | 4.48** | 1.17 | 1.38 |
| Within | 370 | .87 | | .85 | |

*$p < .05$.
**$p < .01$.

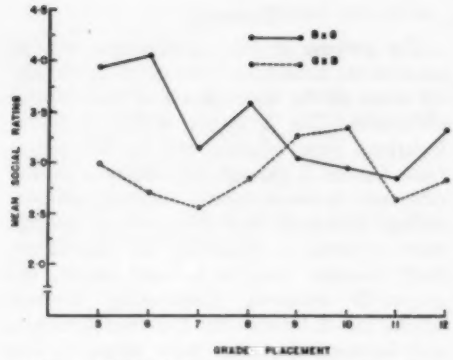


FIG. 1. MEAN OPPOSITE-SEX RATINGS OF BOYS AND GIRLS ON THE NEED-PLAYMIRTH SOCIAL SITUATION

The analyses presented in Table 4 are also relevant to the hypothesis that boys' perceptions of girls are more positive than girls' perceptions of boys. For each need situation the opposite-sex ratings of the boys are significantly higher than the similar ratings of the girls. These results suggest that boys perceive their female classmates as being better able to satisfy their succorance and playmirth needs than the girls perceive them in terms of the same social needs.

## DISCUSSION

This study supports the hypothesis that same-sex social interactions are perceived by preadolescent and adolescent children as more reinforcing than social interactions with the opposite sex. The hypothesis that the difference between boys' and girls' perceptions of their own sex and opposite sex remains constant was upheld in the succorance-need ratings of both sexes. The hypothesis was again upheld in the boys' ratings of their own sex and opposite sex for the playmirth situation. However, the girls' ratings in the playmirth situation showed that after Grade 7, girls perceived boys as increasingly more capable of satisfying their playmirth needs although they still prefer female companions.

There are very few comparable studies of the development of heterosexual affiliations. Harris and Tseng (1957), in a study that is particularly relevant, found that same-sex attitudes for both boys and girls were more favorable than opposite-sex attitudes at all grades. The percentage of favorable attitudes expressed by boys towards girls was found to decrease slightly from three to twelve, though they are generally higher at all grades than the attitudes of girls towards the opposite sex. The percentage of favorable attitudes expressed by girls of boys was found to decrease until Grade 6, then increase until Grade 9, after which they remain constant. From Grade 8 on the degree of favorable attitudes expressed towards the opposite sex for both boys and girls is essentially equal. The similarity of these trends to those for the opposite-sex ratings on the playmirth situation is striking.

These studies appear to support the view that early sex-typed behavior is maintained throughout the preadolescent and adolescent years. The negative expectancies attached to the behavior of opposite-sex peers acquired during the preschool years are maintained because the attitudes and behavior of the opposite sex are not reinforcing. Early expectancies concerning the social behavior of same-sex peers are generally reinforced and appear to maintain themselves throughout the school years and, possibly, into adulthood. The only divergence from this pattern occurs for the girls who display definite changes in their attitudes and perceptions of boys. These changes are probably related to the onset of pubescence and to the concomitant changes in the definition of acceptable female behavior; i.e., after pubescence girls are not only permitted to have social interactions with boys, but they are probably encouraged to do so. At least during the adolescent years girls perceive boys as being as capable of satisfying their playmirth needs as the boys perceive the girls. We might speculate that this state of affairs leads to a measure of social harmony between the sexes which has not existed since the very earliest preschool years.

The over-all decrease in mean ratings from Grades 5 through 12 warrants comment. A trend analysis, reported elsewhere (Meyer, 1957), of the mean ratings for each need situation reveals in almost all cases a significant decelerating trend with increasing grade placement. The interpretation of these grade trends is complex. The most reasonable of

several plausible interpretations can be developed from an examination of the reference population used by the *Ss*. It will be recalled that all *Ss* were required to define a reference continuum using "all people you have ever known" as the basic population. For the younger *Ss* it seems reasonable to believe that a large proportion of all their acquaintances would include their classmates, whereas in the upper grades, classmates constitute a relatively small proportion of all acquaintances. If such an hypothesis is plausible then the observed decreases in mean sociometric ratings probably reflects the social attractiveness of people outside of school for the older *Ss* and explains in part the concomitant decrease in the perceived ability of classmates to satisfy social needs.

The importance of the reference population in the use of these scales cannot be overemphasized. Though the writer can bring no direct evidence to bear on the foregoing interpretation, there is evidence to support a similar interpretation of the significantly lower average ratings made in the need-succorance social situation. Examination of the individuals used by the *Ss* in defining the *most* position on the reference continuum for each social situation shows that the majority of the *Ss* selected an adult for the most preferred position on the succorance scale and a peer on the playmirth scale. When called upon to rate their peers in terms of the former frame of reference it seems unlikely that many classmates would approach the most preferred category. However, when called upon to rate their classmates in terms of the latter frame of reference, there is a greater likelihood of higher ratings.

The conclusions drawn from this study are limited to a rural community and to the social needs of playmirth and succorance. It seems likely that in a more urban community the greater social sophistication of the pupils would have led to different trends. Judgmental situations derived from different social needs and the use of a different reference population would have similarly revealed different relationships from those here described. The influence of these variables on the development of heterosexual affiliations merits further research.


## SUMMARY

The purpose of this investigation was to examine the influence of two social-psychological needs on the development of heterosexual affiliations. The Syracuse Scales of Social Relations were administered to 387 pupils from Grades 5 through 12. Analysis of the differences between same-sex and opposite-sex ratings indicated that the same-sex ratings were significantly higher for the succorance-need situation and, to a lesser degree, the playmirth situation. Contrasting developmental trends between the two need situations and between the sexes were reported. The girls' ratings of the boys on the playmirth situation changed to a positive direction after Grade 7. This shift was not evident for the boys nor was it evident for either sex on the succorance-need situation. The data were interpreted as supporting the notion that early sex-typed behavior is maintained throughout the school years by means of a system of social reinforcements.


## REFERENCES


BONNEY, M. E. Choosing between the sexes on a sociometric instrument. *J. soc. Psychol.*, 1954, **39**, 99–114.

EDWARDS, A. *Manual for the Personal Preference Schedule.* New York: Psychological Corp., 1954.

GARDNER, E. F., & THOMPSON, G. G. *Social relations and morale in small groups.* New York: Appleton-Century-Crofts, 1956.

HARRIS, D. B., & TSENG, S. C. Children's attitudes toward peers and parents as revealed by sentence completions. *Child Develpm.*, 1957, **28**, 401–411.

HATTWICK, L. A. Sex differences in behavior of nursery school children. *Child Develpm.*, 1937, **8**, 343–355.

HORROCKS, J. E., & THOMPSON, G. G. A study of the friendship fluctuations of rural boys and girls. *J. genet. Psychol.*, 1946, **69**, 189–198.

KLETT, C. J. Performance of high school students on the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 68–72.

KOCH, HELEN A. A study of some factors conditioning social distance between the sexes. *J. soc. Psychol.*, 1944, **20**, 79–107.

KUHLEN R. G., & LEE, B. Personality characteristics and social acceptability in adolescence. *J. educ. Psychol.*, 1943, **34**, 321–340.

LINDZEY, G., & BORGATTA, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass: Addison-Wesley, 1954.

MACFARLANE, JEAN, HONZIK, H. P., & DAVIS, M. H. Reputation differences among young school children. *J. educ. Psychol.*, 1937, **28**, 161–175.

MEYER, W. J. Relationships between social need

affiliations. Unpublished doctoral dissertation, Syracuse Univer., 1957.

MORENO, J., *Who shall survive?* Washington: Nervous & Mental Disease, 1934.

MURRAY, H. A. *Explorations in personality: A clinical and experimental study of fifty men of college age.* New York: Oxford Univer. Press, 1938.

RADKE, M. *The relation of parental authority to children's behavior and attitudes.* Minneapolis: Univer. Minnesota Press, 1946.

ROTTER, J. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* New York: Row, Peterson, 1957.

TUDDENHAM, R. D. Studies in reputation: I. Sex and grade differences in school children's evaluations of their peers. II. The diagnosis of social adjustment. *Psychol. Monogr.*, 1952, **66**, No. 1 (Whole No. 333).

WALKER, HELEN, & LEV, J. *Statistical inference.* New York: Holt, 1953.

Received April 10, 1958.

# MOTIVATION AND LANGUAGE BEHAVIOR: A CONTENT ANALYSIS OF SUICIDE NOTES[1]

CHARLES E. OSGOOD AND EVELYN G. WALKER

*Institute of Communications Research, University of Illinois*

WHENEVER a person produces a message, whether it be conversation, an ordinary letter to a relative, or a suicide note, he employs a complex set of encoding habits. It seems reasonable to assume that these language habits are organized in much the same way as the habits underlying nonlanguage behavior and that the general principles of learning and performance therefore apply equivalently in both cases. This paper is concerned with the effects of motivation upon language behavior. It is assumed that the author of a suicide note—presumably written shortly before he takes his own life—is functioning under heightened motivation. Therefore, the structure and content of suicide notes should differ from both ordinary letters and from simulated suicide notes in certain ways predictable from a general theory of behavior. Following a brief theoretical discussion, we describe the application of a number of relevant content measures to a comparison, first, of suicide notes with ordinary letters to relatives and, second, of suicide notes with faked notes. Many of these measures differentiate in predicted ways suicide notes from normal control notes; a smaller number differentiate suicide from simulated suicide notes, suggesting that nonsuicidal individuals are able to adopt the state of the suicidal person in some respects but not in others.

Language habits, like habits in general, appear to be organized into hierarchies of alternatives. We shall assume that increased drive has two distinct effects upon selection within such hierarchies: generalized energizing effects and specific cue effects (cf., Osgood, 1957, for a more complete analysis).

The generalized energizing effects of drives are characterized by a nonspecific facilitation of all habits. Following the views expressed by Hebb (1955), one may identify the generalized energizing effects of drives with arousal of a neural system in the brainstem from which there is diffuse, nonspecific projection into the cortex, these impulses having a summative, "tuning-up" function. Assuming a multiplicative relation between habit strength and drive in producing reaction potential (cf., Spence, Farber, & McFann, 1956), the effect of increasing drive should be to make the dominant alternatives within all hierarchies even more probable relatively. Our first prediction, therefore: (A) *Suicide notes will be characterized by greater stereotypy than messages produced under lower degrees of motivation.* Suicide notes should therefore be more repetitious, less diversified in lexical content, use fewer adjectival and adverbial qualifiers, more familiar words and phrases, and so on. However, since the maximum strengths of habits are assumed to be asymptotic, extreme increase in drive should force many competing habits toward a common maximum and hence produce interference and blocking. Therefore: (B) *If extremely high levels of drive can be assumed, suicide notes should display greater disorganization of language behavior.* This would include various kinds of errors, breaking up of messages into shorter units, and similar phenomena.

To the extent that drive states are accompanied by distinctive sensations—e.g., thirst sensations, feelings of anxiety, sensations of pain—these distinctive cues can become associated with certain alternatives within habit hierarchies through the operation of ordinary learning principles. The presence of such cues, as directive states, will have the effect of modifying the probability structure of behavioral hierarchies, increasing the probability of some alternatives, decreasing the probability of others. This leads to the following prediction: (C) *Suicide notes should be characterized by increased frequency of those grammatical and*

[1] The authors wish to express their thanks to Joseph B. Casagrande of the Social Science Research Council for locating the suicide notes used in this study and to Edwin S. Shneidman of Los Angeles for lending them to us. The notes we used will appear in *The Cry for Help* (McGraw-Hill, 1958), to be edited by Edwin S. Shneidman.

*lexical choices associated with the motives leading to self-destruction.* On a rather mundane level, this means that suicide notes should contain a relatively high frequency of self- and other-critical statements. Less obviously, they should contain a high frequency of what Skinner (1957) calls "mands"—constructions of the demand, command, request type that express needs of the speaker and require some behavior on the part of the listener for their satisfaction. Finally, if two or more motives are operating, and their cues are associated with selection of different alternatives within hierarchies, one may expect oscillation between the responses associated with each state. Since it seems reasonable to assume that suicidal people will often be functioning under competing motives, e.g., self-criticism vs. self-protection, spouse-aggression vs. spouse-affection, etc., we may predict that: (D) *Suicide notes should be characterized by more evidence of conflict than messages produced under nonsuicidal states.* Among indices of conflict would be use of constructions with *but, however, if,* and the like, qualification of verb phrases, and ambivalence in the assertions made about significant persons.

## METHOD

The suicide materials for this study consisted of two samples. The first was a set of 100 genuine suicide notes, 50 written by men and 50 written by women just prior to taking their own lives. These were obtained from Edwin S. Shneidman from his Los Angeles files. For comparison purposes, we obtained a sample of ordinary letters written to 100 members of a panel of *S*s in the Champaign-Urbana area; this panel had been used for other purposes in connection with research on the communication of mental health information. Since many of the quantitative measures we wished to make made it desirable that the messages include at least 100 words or so, the total sample was reduced to the following: 40 male suicide; 29 female suicide; 13 male control; 59 female control. The second set of materials received from Edwin S. Shneidman consisted of 33 paired notes, one of each pair being a genuine suicide note and the other a simulated suicide note; a key to which was which accompanied this set in a sealed envelope. We decided to use this set as a final test of our measures, after trying them out against the known suicide and normal letters. It was expected, however, that certain measures that would discriminate between suicide notes and ordinary letters probably would not do so between genuine and deliberately faked suicide notes—particularly measures reflecting the specific content of the message.

Quantitative measures designed to test the four general predictions—intended indices of *stereotypy,* of *disorganization,* of *directive state,* and of *conflict*—were devised and applied to the samples of known suicide notes and control letters. Sixteen measures were applied, along with certain additional analyses. Some of the measures are standard and well known in content analysis work; others were developed by us for this purpose. These were probably not the best measures that could have been devised, and they certainly do not exhaust the possibilities, but they do represent a considerable variety of quantitative estimates. The two investigators worked together in devising the measures, stabilizing the rules, and applying them to a small sample of notes. Each measure was then applied as consistently as possible to the total materials by one of us, not by both. We therefore have no direct evidence on the reliability of our measures across coders. For some of the measures, the objectivity of what was counted (e.g., number of *repetitions, number of syllables per word*) reduced the seriousness of this problem. Several of the less objective measures (e.g., *evaluative assertion analysis, distress-relief quotient, type/token ratios, cloze procedure*) have been checked for reliability by their authors, and these reports are in our references. To avoid redundancy, the detailed description of the measures will be given in connection with the results obtained with them.

## RESULTS

### *Suidice Notes vs. Ordinary Letters to Friends and Relatives*

The differences between scores of males and females were tested for statistical significance separately within suicide and control groups. If no sex difference was found, the male and female letters within groups were combined and the total suicide vs. control samples were then compared statistically. If a sex difference did appear, separate analyses for differences between suicide notes and controls were made for each sex. Nonparametric tests of significance were used, generally the median test, and occasionally chi square. In the former case, levels of significance were evaluated by reference to the Mainland and Murray tables (1952). Conservative estimates of the significance of the differences are given since a two-sided hypothesis was tested in spite of the fact that the direction of the difference was predicted in all cases.

### *Stereotypy Measures*

1. *Average number of syllables per word.* We would expect a person functioning under high drive to select words in terms of his strongest

habits, i.e., familiar high frequency words. Since, as Zipf (1949) has shown, there is an inverse relation between length of words and their frequency, and since longer, rarer words typically have more syllables, it follows that ordinary letters should have more syllables per word on the average than suicide notes. The total number of syllables per message, as estimated from breath pulses, was divided by the total number of words per message to obtain this index. There were no sex differences on this measure. Differences between suicide notes and control letters did not reach statistical significance but were in the expected direction.

2. *Type/token ratio* (TTR). This measure is obtained by dividing the number of *different* words by the total number of words in each message. It has been shown to be a good index of lexical diversity, differentiating between educational levels, telephone vs. ordinary conversation, and so on (cf. Johnson, Fairbanks, Mann, & Chotlos, 1944). If high drive increases stereotypy, we would expect suicide notes to display lower TTRs than ordinary letters to friends and relatives. There were no sex differences on this measure, and differences between suicides and controls were significant at the .01 level in the predicted direction.

3. *Repetitions*. Another index of stereotypy in messages is redundancy in what is talked about. We would expect people under high drive to repeat phrases more often than people under low drive. Here repetition of single words did not count (cf. TTR above), but phrases and parts of phrases of more than one word did. For example, in "... I really love you very much ... and *I really* do *love you* ...," the part phrase *I really love you* would count as repetition of 4 words. For each message, the number of words repeated in this fashion was divided by the total number of words as an index of repetitiousness. Here, again, there were no differences between sexes, but the difference between suicide notes and ordinary letters was significant at the .01 level.

4. *Noun-verb/adjective-adverb ratio*. This measure—a modification of the familiar verb/adjective ratio (Boder, 1940)—was obtained by dividing the total number of nouns and verbs contained in the message by the total number of adjectives and adverbs. Definition of nouns, verbs, adjectives, and adverbs was done on the basis of whether the words could be substituted in linguistic test frames characteristic of the particular grammatical form. The rationale for the analysis is that under high drive states there should be less tendency toward modification of noun and verb forms, toward discriminative qualification of simple assertions, in line with our assumptions about the generalized energizing effects of drives. The prediction therefore follows that the ratio should be higher in the suicide than in the normal letters. The results bore out this prediction at the .01 level of confidence.

5. *Cloze measures*. Taylor (1956) has devised a method of estimating redundancy or stereotypy in which a message is "mutilated" by substituting a blank for every *n*th word (say, every fifth word, as used here) and *S*s try to fill in these missing items. Presumably, the more predictable the message as a whole, the more accurately *S*s can perform this task and, hence, the higher will be the cloze score. It follows that suicide notes should generate higher cloze scores than control notes. Subsamples of 10 male suicide, 10 male control, 10 female suicide, and 10 female control notes were mutilated by substituting blanks for every fifth word. Because sex differences in content might be significant here, we had 34 male *S*s fill in the male notes and 31 female *S*s fill in the female notes; suicide and control letters were alternated in order of presentation. Each *S*'s mean cloze score for the 10 suicide notes and the 10 control letters was computed. A chi square test was used to determine whether the proportion of *S*s having mean suicide cloze scores higher than their control scores deviated significantly from chance. For male *S*s completing male materials, differences were significant at the .01 level in the expected direction; for females completing female material, however, there were no differences whatsoever.

6. *Allness terms*. People speaking or writing under high drive or emotion could be expected to be more extreme or polarized in their assertions. They should use more terms that permit no exception, e.g., *always, never, forever, no one, no more, everything, everyone, completely, perfectly*, and so on. Strictly speaking, this is not a measure of stereotypy, but it should be affected by generalized drive level. The number of such terms in each message was divided by

total words and expressed as a rate per 100 words. Suicide notes yielded significantly more allness terms (.01 level), and there were no sex differences.

*Disorganization Measures*

1. *Structural disturbances.* Extremely high levels of drive should result in disruption of the myriad of delicately balanced language encoding habits, according to theoretical analysis. To obtain a disturbance measure, the coder took the attitude of an English composition teacher, noting all grammatical, syntactical, spelling, and punctuation errors, and even clearly awkward constructions. Points where material was obviously omitted, e.g., "I don't ( ) him any more," were also counted. The index was the number of such errors expressed as a rate per 100 words. There were no sex differences here and no significant differences between suicide and control notes, although the latter difference was clearly in the expected direction.

2. *Average length of independent segments.* We assume that people encoding under stress will tend to break their utterances into short, explosive units. Here we are interested in sentence length, but must correct for compound sentences joined together by conjunctions like *and* and *but*. The coder divided each message into the number of segments that could stand by themselves as sentences. The index was the total number of words in each message divided by the number of such segments, yielding the average number of words per independent segment. Although there were no sex differences for control letters, there were for suicide notes, male suicides using significantly longer segments (.05 level). Comparing suicides with normals, we find no differences for females but a difference significant at the .05 level for males—male suicide notes used significantly *longer* independent segments than ordinary letters written by males, a finding that is contrary to the direction predicted.

*Directive State Measures*

1. *Distress/Relief Quotient (DRQ).* This well-known measure developed by Dollard and Mowrer (1947) is the ratio of distress-expressing phrases to the sum of these plus relief-expressing phrases, the former being indicative of disturbing drive states and the latter of the reduction of such states. This measure obviously depends to a considerable degree on the judgment of the coder. Here we found definite sex differences for both control and suicide messages; females yielded higher ratios (more distress-expression), in both cases significant at the .05 level. This difference may reflect a trait of masculine reticence in our culture. And, as might be expected from the nature of the suicide situation, both male and female suicide notes displayed higher DRQs than ordinary letters to friends and relatives (.01 level in both cases).

2. *Number of evaluative common-meaning terms.* Common-meaning terms in a language are those upon whose denotation and connotation people must agree if they are to understand one another. Examples would be *sweet, round, table, thunder, run, eat,* and so on. They are in contrast to attitude objects, like *labor union* and *ex-Senator McCarthy,* upon whose connotative meanings, at least, communicators need not agree. *Evaluative* common-meaning terms are those, like *unfair, dangerous, sweetheart,* and *drunkard,* which can be judged as clearly related to either *good* or *bad*. Our index here is simply the total number of such terms in each message divided by the total number of words in the message. There are no differences between sexes for either suicides or normals in the simple *number* of evaluative common-meaning terms (in contrast to the distress/relief measure above and percentage of positive evaluative assertions below), but differences between suicides and normals are significant at the .01 level, with suicides having more evaluative terms.

3. *Positive evaluative assertions.* Evaluative Assertion Analysis has been described in detail elsewhere (Osgood, Saporta, & Nunnally, 1956; Pool, in press). In essence, it involves the linguistic isolation of statements that assert a relation between an attitude object and either another attitude object or an evaluative common-meaning term. Examples: *I* (EGO)/*have always respected/you* (FATHER); *You* (SPOUSE)/*could never stand being/simply a loyal helpmate.* Assertive relations can be either associative (*have always respected*) or dissociative (*could never stand being*). For present purposes, analysis was restricted to those attitude objects

representing the significant persons in Ego's life and their relation to Ego; e.g., EGO, ALTER (person written to when not spouse, parent, or child), SPOUSE, CHILD, MOTHER, FATHER. The index with which we are presently concerned was obtained by dividing the number of positive evaluative assertions by the total number of evaluative assertions, positive and negative, i.e., the proportion of positive evaluations. It would be expected that this measure would correlate highly and negatively with the Distress/Relief Quotient, and it does. Although sex differences were not significant, they were in the direction of greater negative evaluation by females in both suicide and normal letters. Differences between suicide and ordinary letters home were significant at the .01 level and in the expected direction.

4. *Time orientation.* It was expected that the motivational state characteristic of suicide might direct interest of the writer away from the present toward the past. Therefore, suicide notes should contain fewer statements referring to the present and the future but more referring to the past. Examples: present reference—*I love you, I'm afraid that . . .*; past references—*I have tried . . . Everything you've done . . .*; future reference—*. . . who will always love you, Tell my parents. . . .* We measured both the proportion of total references which were to present time and the imbalance of nonpresent references toward past vs. future. Contrary to our expectations, there were neither significant differences between sexes nor between suicides and controls.

5. *Mands.* According to Skinner (1957), a *mand* is an utterance which (*a*) expresses a need of the speaker and which (*b*) requires some reaction from another person for its satisfaction. It is usually expressed in the form of an imperative, where the verb comes early in the utterance (e.g., *Don't feel too bad about this* or *Please understand me*), but is not restricted to this form (e.g., *I wish I could see you, I hope you understand*, or *May God forgive me*). Our index was the number of such constructions, expressed as a rate per hundred words. This proved to be one of the most useful measures in our arsenal. In the present test situation, differences between sexes were not significant, but differences between suicide notes and ordinary letters to friends and relatives were significant at the .01 level in the predicted direction.

*Conflict Measures*

1. *Qualification of verb phrases.* When a speaker or writer is in conflict about the topics being discussed, it seems likely that he will modify or qualify his assertions away from the flat, direct present or past tense, e.g., from *I was good to you* to something like *I used to be good to you*, or *I tried to be good to you.* To quantify this characteristic of messages, the coder first bracketed each complete verb phrase for which a single verb could be substituted, e.g., for *I* (*could have helped*) *you more* we can substitute *I* (*loved*) *you more*, where the one word *loved* substitutes structurally for the three words *could have helped*; then the coder totaled the number of words in these brackets and divided by the number of such brackets. The larger this ratio, the greater the amount of excess, qualifying material. There were no sex differences on this measure, but differences between suicide and normal letters were significant at the .01 level in the expected direction.

2. *Ambivalence constructions.* There are a number of syntactical constructions in English that may directly express ambivalence, conflict, and doubt on the part of the speaker: *but, if, would, should, because* (*for, since*), *well, however, maybe, probably, possibly, seems, appears, guess, surely, really, except*, etc. Certain question forms also express the same indecisive state, e.g., *Must I do it? Why do I try at all?* The coder determined the number of such forms in each message and expressed it as a rate per 100 words. Differences between sexes were not significant; differences between suicide notes and ordinary letters were significant at the .01 level.

3. *Percentage of ambivalent evaluative assertions.* The essential nature of evaluative assertion analysis has already been described (see Directive State Measures numbered 2 and 3). If a speaker displays perfect consistency or lack of ambivalence, then all of the assertions relating to each attitude object or association between each pair of attitude objects will have the same sign. For example, assertions concerning the self would be either consistently positive or consistently negative—*I am no*

*good; I have been a failure; Luck has not been on my side;* or for the relation of Ego to Spouse, *I have always loved you; You relied on me; I tried to help you; My Darling Wife.* Ambivalence, on the other hand, is indicated by assertions of different signs in the same set, e.g., *I love you, Honey; You never trusted me; I quarreled with you; You stuck by our marriage, though.* Our index of assertion ambivalence was the total number of deviant assertions (i.e., the number of least frequent signs in each set, summing over sets) expressed as a proportion of total assertions. For this conflict measure, also, there were no sex differences, but suicide vs. control differences were significant at the .01 level and in the predicted direction.

*Genuine vs. Simulated Suicide Notes*

We had originally planned to apply the measures that had successfully differentiated suicide from control letters blindly to the entire set of 33 paired suicide and simulated notes. Unfortunately only 13 of these pairs included both suicide and faked notes of sufficient length to make most of our measures meaningful. In attempting to predict which of these 13 pairs were the genuine suicide notes, we eliminated those measures which had failed to differentiate suicides from normal controls (*structural disturbances, average length of independent segments, time orientation*), those which obviously and nonsubtly reflected the suicide topic and hence would be readily faked (*distress-relief quotient, evaluative terms, positive evaluative assertions*), and the *cloze procedure* (too few messages of sufficient length). A prediction of suicide vs. simulated was made for each of the 13 paired notes on the basis of each of the remaining nine measures, and the final prediction was based on which note in each pair garnered the most suicide votes. Our quantitative predictions proved to be correct in 10 out of 13 cases, a value significant at the .05 level.

However, before checking the accuracy of these quantitative predictions against the key in the sealed envelope, both authors independently assigned all 33 pairs to suicide or fake categories on an intuitive basis. One of us got 31/33 correct and the other 26/33. To check on the possibility that we had actually been employing cues derived from our previous quantitative coding, we had eight graduate students with no prior experience with these notes assign the pairs to genuine suicide and fake categories. They were successful on the average in 16.5/33 cases, exactly chance. So it would appear that familiarity with a large sample of known suicide vs. nonsuicide notes, or sensitivity to cues derived from quantitative measures, or both, contributes to successful identification of genuine suicide notes.

Knowing which of the 33 pairs are genuine suicide notes, we may now ask which of our quantitative measures, successful in differentiating suicide from ordinary letters home, are also successful in differentiating genuine suicide from pseudo notes. This analysis should indicate which encoding characteristics of the suicidal individual can be intuited and, hence, faked by the nonsuicidal person, and which cannot. We may look first at the very small sample of 13 pairs where both genuine and pseudo notes could be coded.

*Stereotypy Measures:* Of these measures, three (*syllables per word, repetitions,* and *allness terms*) were clearly in the expected direction but not significantly so. One, *noun-verb/adjective-adverb ratio* was significant at the .05 level in the predicted direction.

*Directive State Measures:* Of the directive state measures, *DRQ, frequency of evaluative terms,* and *proportion of positive evaluative assertions* did not differentiate (as expected), but *mands* did differentiate significantly at the .05 level.

*Conflict Measures:* Of the three conflict measures, one was not significant (*qualification of verb phrases*), one was significant at the .05 level, but in *the wrong direction* (*ambivalence constructions*) and one was barely significant in the predicted direction at the .10 level (*proportion of ambivalent assertions*).

If we enlarge our sample to 24 suicide and 18 faked notes by scoring all notes of sufficient length, regardless of their pairing, about the same results appear: *Stereotypy measures* tend in the right direction, but only the *noun-verb/adjective-adverb ratio* significantly so; *mands* just miss significance at the .05 level; *proportion of ambivalent assertions* approaches significance in the expected direction, but the other *conflict measures* are either nondifferential

TABLE 1
WORDS INCLUDED IN FIVE OR MORE OF THE GENUINE AND SIMULATED SUICIDE NOTES

| Nouns | | Verbs | | Adjectives and Adverbs | |
|---|---|---|---|---|---|
| Genuine | Simulated | Genuine | Simulated | Genuine | Simulated |
| everything (9) | life (13) | love (19) | know (10) | good (15) | good (9) |
| way (out) (9) | way (11) | tell (13) | leave (9) | sorry (11) | sorry (8) |
| wife (9) | way (out) (8) | know (12) | think (9) | only (7) | happy (6) |
| love (8) | thing (8) | hope (11) | have (8) | dear (6) | all (5) |
| mother (8) | wife (7) | please (11) | please (8) | bad (5) | |
| thing (8) | all (8) | think (11) | love (7) | | |
| God (6) | love (5) | do (10) | forgive (6) | | |
| time (s) (6) | insurance (5) | get (9) | hope (5) | | |
| darling (5) | | say (9) | seem (5) | | |
| dear (5) | | take (9) | see (5) | | |
| honey (5) | | give (8) | tell (5) | | |
| life (5) | | want (8) | | | |
| one (5) | | feel (7) | | | |
| person (5) | | goodbye (7) | | | |
| something (5) | | have (7) | | | |
| trouble (5) | | make (7) | | | |
| way (5) | | go (6) | | | |
| year(s) (5) | | help (6) | | | |
| | | see (6) | | | |
| | | forgive (5) | | | |
| | | try (5) | | | |
| | | take care of (5) | | | |

or significant in the wrong direction. Interestingly, the two disorganization measures, which were computed for this larger sample, approach (*structural disturbances*) or reach (*length of independent segments*) significance at the .05 level and in the predicted direction.

Although gross measures of "what is talked about" like the DRQ and negative evaluative assertions may not differentiate genuine from pseudo suicide notes, we may ask if a more detailed content analysis might reveal differences. Accordingly, the frequency with which lexical words (nouns, verbs, adjectives, and adverbs) were used in the 33 genuine and facsimile suicide notes was analyzed. Since the sample of *S*s was small, and hence liable to bias by discussion of a particular topic by a single *S*, subject-frequencies rather than word-frequencies per se were counted, i.e., the use of a given word by a given *S* was only counted once no matter how often he used it.

Table 1 gives the words included in five or more of the 33 suicide notes and in five or more of the simulated notes. Again, we note evidence for greater stereotopy in the suicide group; in every category, suicide notes display a greater sharing of common lexical items than do simulated notes. Some of the differences in choice of most frequently used lexical items are interesting: Suicide notes are more heavily loaded with terms of endearment (*darling, dear, honey*) and references to *mother*, whereas faked notes have more abstractions (*life, way, all*) and references to *insurance*. Whereas genuine suicide notes are replete with verbs referring to simple action (*tell, do, get, say, take, give*), faked suicide notes include relatively more verbs referring to mental states (*know, think, seem, see*). The genuine suicides have more stress on positive states (*love* 19 and *hope* 12 vs. 7 and 5 for the same words in pseudo notes), whereas the simulated notes have 9 references to *leave* vs. only 3 for suicides. To summarize this rough content comparison, genuine suicide notes reflect ambivalence toward loved ones through higher frequency of *positive evaluative terms* (cf., *ambivalent assertions* above) and they reflect greater *concreteness*.

Finally, a contingency analysis[2] of some of the major content categories in suicide and pseudocide notes was made in an attempt to get at the association structures characteristic of the two groups. The content categories given in Table 2 were used. Before discussing the results of the contingency analysis, some of the

[2] This method is described in some detail in Pool (in press).

TABLE 2
RELATIVE FREQUENCY OF CASES REFLECTING VARIOUS CONTENT CATEGORIES

| Categories | Genuine | Simulated |
|---|---|---|
| Spouse praise, defense, love | .69 | .42 |
| Self criticism | .48 | .39 |
| I'm sorry; forgive me | .45 | .36 |
| Self praise, defense | .39 | .06 |
| Children | .36 | .33 |
| Goodbye, farewell, etc. | .30 | .12 |
| Feelings of confusion, being tired, etc. | .27 | .42 |
| Spouse criticism | .24 | .03 |
| "Way out" | .24 | .39 |
| Physical disabilities, symptoms | .21 | .18 |
| Parents | .21 | .06 |
| God and religion | .21 | .06 |
| Material possessions | .21 | .00 |
| Reference to suicidal act | .18 | .33 |
| Money, bills, debts | .15 | .06 |
| Notify, tell someone | .15 | .03 |
| Isolation, loneliness | .15 | .00 |
| Insurance, etc. | .12 | .21 |
| Reference to suicide note | .12 | .06 |
| Instructions about own remains | .12 | .03 |
| Job | .09 | .03 |
| Love triangles (other man, woman) | .09 | .06 |
| "Fate," "Life," "World," etc. | .03 | .33 |
| Sex relations | .03 | .00 |

differences in relative frequency of reference to these categories in the genuine and spurious suicide notes are worth noting: Suicide notes have relatively higher frequencies of reference to *self praise and defense*, *goodbyes* and *farewells*, *criticism of the spouse*, references to *parents*, *God and religion*, and *material possessions*; simulated notes refer relatively more frequently to *feelings of confusion*, *being tired*, and the like, to the *suicidal act* itself, to *insurance*, and to abstractions like *Fate*, *Life*, and *The World*.

Expected and obtained contingencies among these categories (for genuine suicide and simulated notes separately) were obtained in the following way: Letting A and B represent two content categories, the *expected contingency* is $p_{AB}$, i.e., the probability of both A and B being present in notes of a given type, based on their separate rates of occurrence. The *obtained contingency* is simply the relative frequency of actual co-occurrence, i.e., the percentage of notes of a given type in which contents A and B are actually both present. The obtained contingency may be either greater than (association) or less than (dissociation) the expected or chance contingencies. Significances of deviations from chance expectancies are estimated in terms of the standard error of the expected percentage. Because of the crude nature of this analysis and the rather small $N$, significances at the .10 level or better were used as the basis for the following summary statements; they should be considered to be merely suggestive.

In the genuine suicide notes we find *criticism of the spouse* associated with references to *insurance*, *money*, *bills and debts*, and *requests to notify* someone of his death. As might be expected, *requests to notify* are associated with references to the *suicide note* itself and with *instructions about handling one's remains*. Expressions of feeling *isolated and lonesome* are associated with references to *money*, *bills and debts*, and to *love triangles*. References to the *parents* appear with statements about taking a *"way out"* and with references to *material possessions*. References to *own children* appear with *instructions about handling remains*. Again as would be expected, references to *money*, *bills and debts* co-occur with references to *material possessions*. Less obviously, references to *the job* are contingent upon references to the *suicide note* itself; *self praise* is contingent upon references to *insurance*.

In the simulated notes, references to *own children* are associated with references to *God and religion*, while references to *parents* are associated with stereotyped abstractions, *Fate*, *Life*, *the World*. When people write "make-believe" suicide notes, references to *the suicide act* itself tend to be accompanied by references to *God and religion*, and talking about a *"way out"* appears with comments about *money*, *bills and debts*. Expressions of *sorrow*, *regret*, *and asking for forgiveness* appear with saying *goodbye* and *farewell* in these faked notes. There are also two significant dissociations (co-occurrence *less* than chance at the .10 level)—faked notes that speak of *physical disabilities* do not express *feelings of confusion and being tired*, and the notes which refer to *insurance* do not include expressions of *sorrow*, *regret*, *and asking for forgiveness*.

In viewing the total evidence on association structures, one gets the following general impression: When people produce fake suicide notes "on demand," they generally embroider a few standard themes available in our folklore—taking a "way out" of financial and other

problems, asking forgiveness and saying "farewell," pondering on the moral and religious implications of taking one's own life, and so on. The patterns of association in genuine suicide notes suggest more mundane connections—for example, criticism of the spouse being connected in the suicidal person's mind with his financial problems, with being insured and the like, or references to being insured being coupled with self-praise.

## DISCUSSION

Of the four general predictions about the effects of heightened motivation upon encoding, three are borne out clearly in the comparison of suicide notes with ordinary letters to friends and relatives. Suicide notes display greater *stereotypy*—the writer of a suicide note tends to use shorter, simpler words, his vocabulary is less diversified, he is more repetitious, he uses more simple action expressions (nouns and verbs) and fewer discriminative qualifiers (adjectives and adverbs), and his messages are more easily filled in (*cloze* procedure) by others. He also uses more polarized "allness" terms. The effects of the suicidal *directive state* are also clearly evident—in higher distress-relief quotients, in the greater frequency of evaluative terms, and in the smaller proportion of evaluative assertions that are positive in direction. Suicide notes also display the demanding, commanding, pleading nature of this state by higher frequency of *mands*. Suicide notes yield evidence of greater *conflict* of motives—by greater qualification of verb phrases, more ambivalence constructions, and a larger percentage of evaluative assertions about Ego and significant others that are ambivalent in sign. Most of these differences were significant at the .01 level, and they substantiate our general hypotheses about the effects of heightened drive level upon language encoding.

One major prediction was not borne out: there was no evidence for greater *disorganization* of encoding behavior in suicide notes as compared with ordinary letters to friends and relatives. We conclude that the suicide state, at least at the time a note is penned, does not represent a sufficiently high degree of motivation to cause disruption of language skills, but the negative result could also indicate that our measures of disorganization were inadequate or that the hypothesis was wrong. The failure of the time orientation measure to yield any differences may also mean that our notions were wrong. It is also possible that this measure was confounded with that for *mands*; mands usually have future reference and are significantly more frequent for suicide notes.

The comparison of genuine suicide notes with simulated suicide notes, matched for age, education, and general social status, can be considered, on the one hand, a more stringent test of the hypotheses or, on the other hand, an indication of the degree to which a nonsuicidal person can intuit and adopt the encoding content and style of the suicidal person. From the former point of view, we would have to conclude that most of our measures fail to distinguish significantly between the genuine and pseudo notes (excepting the *noun-verb/adjective-adverb ratio*, *mands*, *length of independent segments*, and perhaps the *proportion of ambivalent assertions*). Nevertheless, the quantitative indices that differentiated suicide from ordinary letters, and which might be expected to differentiate genuine from faked notes, did so for 10 out of the 13 matched pairs to which they could be applied.

How well can nonsuicidal writers adopt the encoding content and style of the suicide state? First, nonsuicidal people can obviously intuit the superficial *content* of suicide notes—the distress-expression, the use of evaluative terms, and the decrease in positive evaluative assertions. Less superficially, however, we note interesting differences in the words used (more positively toned terms, expressing the ambivalence of the true suicidal state, and more concrete terms generally) and in the contingencies among content categories (less stereotyped, "story-book" associations in the true suicide cases). Second, although the over-all reduction in significance of differences, as compared with the suicide vs. control analysis, shows that the *style* of the suicidal person can be adopted to some degree by a person merely instructed to write such a note, there are certain exceptions. The person faking a suicide note fails to reflect the demanding, commanding, pleading style (*mands*), the reduced qualification (*noun-verb/adjective-adverb ratio*), and the evaluative ambivalence toward self and others of the genuine suicide notes.

One criticism that could be leveled at this study is that other determinants than motivation might be responsible for the results. It is known that many of the indices used here as tests of motivational effects can be affected by other source characteristics as well. For example, stereotypy measures like length of words and TTR are influenced by the education and IQ level of the source. Our control sample of ordinary letters to friends and relatives could be matched with the suicide notes in terms of sex and age, but that was all. Could the differences we found be accounted for simply on the basis that our suicide note writers were less intelligent and/or less well educated? If we explain the differences in stereotypy in this way, we are unable to explain why the same notes showed no differences in structural disturbances (ordinary English composition, for the most part) and, in fact, for males showed *longer* integrated sentence segments. Furthermore, this would not explain the directive-state differences (e.g., in *mands*). Also, in the genuine-pseudo comparison, where these factors were controlled by matching, differences for the most part were in the same direction, although not as large.

## SUMMARY

Theoretical analysis of the effects of motivation level upon language encoding led to several hypotheses. Messages produced under heightened drive level should (*a*) be more stereotyped, (*b*) be more disorganized, if the motivation level is extremely high, (*c*) reflect the specific nature of the motives operating, and (*d*) reflect conflict of responses if two or more competing motives are operating. These hypotheses were tested by (*a*) a comparison of suicide notes with ordinary letters to friends and relatives and (*b*) a comparison of genuine suicide notes with simulated suicide notes, written by nonsuicidal people. In the first comparison, all of the hypotheses were clearly borne out except that concerning disorganization of encoding skills. In the second comparison, differences were smaller, only certain measures, the *noun-verb/adjective-adverb ratio*, Skinner's *mands*, *length of sentence segments*, and *proportion of ambivalent evaluative assertions* still discriminating significantly. Implications of these result for psycholinguistic theory and for stylistics are considered in the discussion, along with certain criticisms that could be made of the study.

## REFERENCES

BODER, D. P. The adjective-verb quotient. A contribution to the psychology of language. *Psychol. Rec.*, 1940, **3**, 309–343.

DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3–32.

HEBB, D. O. Drives and the C. N. S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

JOHNSON, W., FAIRBANKS, HELEN, MANN, MARY BACHMAN, & CHOTLOS, J. W. Studies in language behavior. *Psychol. Monogr.*, 1944, **56**, No. 2 (Whole No. 255).

MAINLAND, D., & MURRAY, I. M. Tables for use in fourfold contingency tests. *Science*, 1952, **116**, 591–594.

OSGOOD, C. E. *Motivational dynamics of language behavior*. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1957.

OSGOOD, C. E., SAPORTA, S., & NUNNALLY, J. C. Evaluative assertion analysis. *Litera*, 1956, **3**, 47–102.

POOL, I. (Ed.) *Contemporary trends in content analysis*. Urbana: Univer. Illinois Press, in press.

SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

TAYLOR, W. L. Recent developments in the use of "Cloze Procedure." *Journalism Quart.*, 1956, **33**, No. 1.

ZIPF, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.

Received April 10, 1958.

# GROUP VERSUS INDIVIDUAL PROBLEM-SOLVING[1]

WILLIAM L. FAUST

*Pomona College*

Is THE performance of a group superior to the performance of an individual? Studies by Shaw (1932), Watson (1928), and Taylor and Faust (1952) have found that group performance is significantly superior. In the Taylor-Faust study (1952), for example, groups composed of two and four members were found to solve more Twenty Questions problems in fewer questions and in shorter time than did persons working individually.

These studies have based their statistical analysis upon the null hypothesis that no difference exists between group and individual performance except that produced by chance. However, the results indicating that group performance is superior to individual performance have often been interpreted as due to some "group effect" in which individual members of the group are influenced by other members to work on problems along lines of thought which they would not have generated themselves. Marquart (1955) has criticized the validity of such an interpretation in which a person working alone is counted as equal to one group and in which no allowance is made for the fact that a group solution might be the result of any one of the members (perhaps the ablest) rather than the cooperative effort of the group as a whole.

A different null hypothesis and one more appropriate for testing the influence of some group effects can be derived from the discrete probability theory.[2] If, in a certain population, $P_I$ is the percentage of the individuals working alone who can solve a particular problem, then $Q_I$ will be the percentage who fail to solve that problem. The probability of drawing, at random from that population, a single individual who will fail that problem is $Q_I$; the probability of drawing two individuals who will fail is $Q_I^2$; the probability of drawing $m$ individuals, all of whom will fail the problem, is $Q_I^m$. At the same time, the probability that *at least one* of the above $m$ individuals will solve the problem is $1 - Q_I^m$.

Suppose that many individuals have worked separately on a problem and $Q_I$ (percentage) have failed to solve the problem. If groups composed of $m$ members are given the problem, the percentage, $P_G$, of such groups which will solve the problem (and which therefore have at least one member who can solve the problem) is given by

$$P_G = 1 - Q_I^m$$

It should be noted that $P_G$ will be a larger portion than $P_I$ (except when $P_I$ equals either 0.00 or 1.00 in which case they will be equal).

Unfortunately, this latter null hypothesis has one serious limitation: The results from more than one problem cannot be combined unless the problems are homogeneous in difficulty, a condition that rarely exists. This limitation results from the fact that no method is known for testing the significance of the combined results. Since results from separate problems attempted cannot be combined, and since performance of *S*s on a single problem is relatively unreliable, a large number of *S*s would be required to provide an adequate test of the null hypothesis for even a single problem.

Because of this limitation, the two experiments reported here employed an experimental

[1] This study is based upon a dissertation submitted in 1954 to the Department of Psychology and the Committee on Graduate Study of Stanford University in partial fulfillment of the requirement for the degree of doctor of philosophy. This research was done under Project NR 150-149 supported by Contract N6onr between Stanford University and the Office of Naval Research. Permission is granted to the United States Government for the reproduction, translation, publication, use, and disposal of this article in whole or in part. Work on the contract was under the direction of Donald W. Taylor. The author wishes to express his sincere thanks to Donald W. Taylor for his guidance and encouragement, and also to Ramon J. Rhine who assisted the author by conducting the actual sessions for Experiment B at Stanford.

[2] The following model has been rephrased from the presentation by Taylor and McNemar (1955) and Taylor (1954) and by Faust (1954). Irving Lorge and Herbert Solomon have also formulated and employed this model (1955).

design that makes possible the comparison of actual group performance on a number of different problems, with the performance to be *expected* of a group on these problems. This design was devised and suggested to the present writer by Taylor and later described by him in articles (Taylor & McNemar, 1955; Taylor, 1954) published since the completion of the present experiments. The performance to be *expected* of a group was derived in the following manner: *S*s were assigned at random to work either individually or in groups. After the experiment was completed, those who had actually worked alone were arranged by a random procedure into so-called "nominal" groups of the same size as the real groups. The performance of these nominal groups was then scored as though the individuals had actually worked together—i.e., if any individual in the nominal group had solved the problem, the group was scored as having solved it.

Marquart (1955) has used a similar design. However, she had her *S*s work both individually and in groups of three. She then pooled the individual performances for each threesome who at another time worked together in a group and scored such nominal groups in the method described above. In both Marquart's own data and her reanalysis of Shaw's (1932) data the difference between real and nominal groups was not significant although any advantage seemed to favor nominal groups. Marquart's method poses the difficulty of counterbalancing the effects of improvement in solving problems when the same *S*s work many problems both individually and in groups.

These latter two models do not assume either homogeneity of *S*s or homogeneity of problems. The performance of nominal groups scored in the way proposed provides a measure of the performance to be expected from a given size of group on a probability basis—i.e., on the assumption that there is no effect of working in the group, either positive or negative.

In order to use nominal group performance as a baseline against which to compare real group performance, the assumption must be made that if one, or more than one, person in the group solves the problem he will be able to convince the others that the solution is correct. It is possible that the group effect may stimulate a solution in a member who might have failed if he were working alone; however, this person may not be able to persuade the others, and the group as a whole may fail to decide on the correct answer. In order to avoid this confusion, problems were chosen with simple, unequivocal answers so that when a person discovered the answer he could quickly demonstrate that solution and convince others.

## PROCEDURE

The two experiments reported were conducted in different academic institutions—Pomona College (Experiment A) and Stanford University (Experiment B). Both experiments used students from the introductory course in psychology. In Experiment A, 101 *S*s were randomly assigned to work in one of the 15 groups of four members, or to be one of the 41 individuals working alone. In Experiment B, 135 *S*s were randomly assigned to work in 17 groups of four, or to be one of the 67 individuals working alone.

The two experiments were essentially replications of the same design. In Experiment A, 40% of the *S*s were women and 60% men; the corresponding proportions in Experiment B were 75% women and 25% men. Although there was a difference *between* Experiments A and B in the percentage of men and women, *within* each experiment there was *no* difference between real and nominal groups in the percentage of men and women.[3]

Each individual and group of four worked for a one-hour session on seven problems.[4] Four of the problems were spatial problems in which the problem was presented in a diagram along with a verbal description of the problem. The *S* was directed to "draw" the solution.

As examples of the kind of spatial problems used, two will be cited: Katona's match problem (1940, p. 58)—sixteen matches are laid out making five squares and three are to be moved to different positions to make four squares; the nine dot problem (Ruch, 1958, p. 357)—without lifting the pencil from the paper, connect all nine dots by drawing four straight lines.

The other three problems were anagram problems. Each of these problems was a sentence composed of seven words. The seven words were in their correct sentence-order, but the letters within each word were in jumbled order. As the words were solved in order, the words began to form a sentence, and the emergent meaning aided in the solution of the following words.

The four spatial problems and answer sheets for the three anagram problems were contained in a booklet. The booklets were distributed to *S*s along with envelopes containing the anagram words. These anagram problems were held in three large envelopes, one envelope for each one of the three anagram problems.

[3] For a more complete description of the procedure and analysis of the two experiments see Faust (1954).

[4] The problems used can be found in Faust (1954).

Each one of these large envelopes contained seven small envelopes which held the 5 x 7 cards on which the anagram words were typed. This system of envelopes within envelopes required that *S*s open and solve the anagram in each envelope before the next envelope could be opened; therefore, the four *S*s of the group were forced to work jointly on the same step of the multiple-step problem. This system was particularly important for testing groups where the groups would have a decided advantage if they did not work on each word together, but instead each member of the group worked on a different word.

Since groups had to be given specific instructions directing their "group" attack on the problems, two different sets of instructions were used. The instructions for individuals and the instructions for groups were appended to booklets containing the seven problems. The instructions to individuals explained that the purpose of the experiment was to see how efficiency in problem-solving depends upon size of group. The individuals were told that they would have exactly five minutes for each problem, that they probably would not solve all of the problems in the time limit, but that they were not to go back and work on problems after the time limit.

Instructions to groups were similar to those given to individuals, but, in addition, it was explained that the four members were to cooperate, talk freely, and organize their efforts in whatever way seemed most efficient for attacking and solving each problem. However, the four members of the group were directed to work together, simultaneously, on the same problem, or on the same step in multiple-step problems.

TABLE 1

MEAN NUMBER OF SPATIAL AND VERBAL PROBLEMS SOLVED BY INDIVIDUALS AND REAL GROUPS

| | Individuals | Real Groups | Diff. | $p$* |
|---|---|---|---|---|
| Experiment A | | | | |
| Spatial | .80 | 2.87 | 2.07 | .001 |
| Verbal | .68 | 2.67 | 1.99 | .001 |
| Experiment B | | | | |
| Spatial | .58 | 1.82 | 1.24 | .01 |
| Verbal | .94 | 2.53 | 1.59 | .001 |

* The $t$ for uncorrelated series, McNemar (1955, p. 109), was used.

TABLE 2

MEAN NUMBER OF SPATIAL AND VERBAL PROBLEMS SOLVED BY REAL AND NOMINAL GROUPS

| | Real Groups | Nominal Groups | Diff. | $p$** |
|---|---|---|---|---|
| Experiment A | | | | |
| Spatial | 2.87 | 2.20 | .67 | * |
| Verbal | 2.67 | 1.90 | .77 | .02 |
| Experiment B | | | | |
| Spatial | 1.82 | 1.88 | −.06 | * |
| Verbal | 2.53 | 2.25 | .28 | (.10) |

* Not significant at .05.

** The $t$ for uncorrelated series, McNemar (1955, p. 109), was used.

## RESULTS

In view of the different percentage of men and women in the two studies, the data from the two studies will not be combined. A study by Sweeney (1953) provides evidence that men are more likely than women to solve the kind of spatial problems employed here. In addition, it is very likely that women are better on the verbal problems than men. Comparison of the real and nominal groups in Experiment A with those in Experiment B bear out the fact that the data are not entirely comparable.

### *Group Versus Individual Performance*

First, the data will be analyzed using the null hypothesis that no differences, other than chance differences, are to be expected between group and individual performance. The results of the present study support the findings of previous studies that mean performance of groups, at least within the conditions investigated, is superior to the mean performance of individuals working alone. As shown in Table 1, the mean number of problems solved by the groups of four was significantly greater than the mean number solved by individuals for both verbal and spatial problems in both Experiment A and Experiment B.

### *Real Versus Nominal Groups*

The second analysis comparing the performance of real and nominal groups follows logically from two assumptions: (*a*) real groups ought to do better than individuals just on a probability basis; and (*b*) there is no effect of working in a group either positive or negative.

The use of nominal groups and the method of scoring these groups are the crux of the design being used in this study. Working with the two experiments separately, nominal groups were constructed by assigning a number to each of the individuals who worked alone. Cards containing the numbers were then randomized and sets of four individuals were randomly selected.[5] Each of the seven problems

[5] The suggestion has been made that it would be possible to compute the mean performance for all possible combinations of groups of four which could be constructed from the sample of individuals who worked

was recorded separately, and a nominal group was scored as having solved a problem if any one of the members of that group had solved it. A total score for spatial problems and a total score for verbal problems—in each case indicating how many had been solved—was computed for nominal and for real groups.

The results obtained in the comparison of real and nominal groups are summarized in Table 2. In neither Experiment A nor B did the performance of the real groups differ significantly from that of the nominal groups on spatial problems. In contrast, the mean number of verbal problems solved for real groups was significantly larger in Experiment A, though not in Experiment B.

## DISCUSSION

The present study makes two kinds of comparisons: group performance with individual performance and group performance with nominal group performance. On spatial problems, group performance was significantly superior to individual performance in both Experiments A and B. These results are consistent with those of previous studies comparing group with individual performance. However, comparison of the performance of real and nominal groups showed that group performance on spatial problems was no better than would be expected on a probability basis. There was no effect of working in a group, either positive or negative.

On verbal problems, group performance was also superior to individual performance in both Experiments A and B. But the results differ from those for spatial problems in that the performance of real groups was superior to that of nominal groups for Experiment A, but not for Experiment B.

There is an interesting difference between the verbal and spatial problems used besides that implied by their names. The spatial problems are essentially one-step-insight problems, while the verbal problems are multiple-step problems. Lorge and Solomon (1955) have suggested that for multiple-step problems a more complex model is required (Model B, p. 143). In solving multiple-step problems, the group may have one member who can solve one step of the problem, another member who can solve a different step in the problem, so that the group may solve the problem even though no single member of the group could solve that problem alone. This combining of part solutions may account for the significantly superior performance of real groups on the multiple-step verbal problems.

The problem of group performance versus individual performance is one that has both theoretical and practical interest. The present day use of committees to direct the activities of most organizations, the educational emphasis upon the group approach to learning, and the frequent recourse in science to conferences and group research all indicate that, presumably, groups can operate more effectively than individuals. The present study has restricted its investigation to only one aspect of group effectiveness—performance in solving certain kinds of problems. However, working in groups may serve other functions. Participation in groups may be important because it motivates individuals in situations in which motivation may otherwise be inadequate. Group decision-making may be desirable, not because the decision reached is better than that which would be made by individuals, but because the decision will be accepted by the members of the group once it is made. The possible importance of these and other benefits of working in groups should not be minimized. On the other hand, the frequent assumption that groups are superior to individuals in solving problems because of some kind of interaction in the group should be seriously questioned.

The extent to which the conclusion obtained in these experiments can be generalized is, of course, limited by the kinds of problems employed, the characteristics of *S*s sampled, etc. To answer the question as to what kinds of problems, what kinds of individuals, and under what conditions group performance is superior to that expected on a probability basis will require literally many years of investigation. The

alone. Although this could be done, no method is known to the writer for testing the significance of the difference between such a computed mean and the mean performance of the real groups. Consultation with two members of the staff of the Laboratory of Applied Mathematics and Statistics at Stanford suggests that no appropriate method exists. The method used here of constructing nominal groups theoretically does not use all the information available; however, it does permit an appropriate statistical test of obtained differences in performance between nominal and real groups.

two experiments reported make only a small contribution to that answer.

## SUMMARY

The two experiments reported were replications of the same design differing in the percentage of men and women used. There were 15 groups of four and 41 individuals in Experiment A, and 17 groups of four and 67 individuals in Experiment B. All individuals and groups were given four spatial and three verbal anagram problems.

1. This study found, as had previous studies, that group performance was superior to individual performance; groups in both of the two experiments solved significantly more spatial as well as verbal problems than did individuals.

2. When the performance of real groups on spatial problems was compared with the performance of nominal groups, there was no significant difference in either Experiment A or B. Real groups solved no more problems than would be expected on a probability basis.

3. In Experiment A, real groups solved significantly more verbal problems than did nominal groups. In Experiment B, the difference was in the same direction but did not reach significance ($p = .10$).

4. The present research has used a method of establishing a baseline for evaluating group performance. This baseline consists of the performance of nominal groups in which *S*s who worked individually are allotted to form groups of the same size as real groups. The performance of these nominal groups is scored as though the individuals had actually worked together. Methodologically, this design isolates one of the factors of group superiority, namely that of having more persons working on the problem.

## REFERENCES

FAUST, W. L. Determinants of individual improvement and of group performance in solving certain types of verbal and spatial problems. Unpublished doctoral dissertation, Stanford Univer., 1954.

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

LORGE, I., & SOLOMON, H. Two models of group behavior in the solution of eureka-type problems. *Psychometrika*, 1955, **20**, 139–148.

MCNEMAR, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

MARQUART, DOROTHY IRENE. Group problem solving. *J. soc. Psychol.*, 1955, **41**, 103–113.

RUCH, F. L. *Psychology and life.* (5th ed.) Chicago: Scott, Foresman, 1958.

SHAW, M. E. A comparison of individuals and small groups in rational solution of complex problems. *Amer. J. Psychol.*, 1932, **54**, 491–504.

SWEENEY, E. J. Sex differences in problem solving. Stanford: Dept. Psychol., Stanford Univer., 1953. (Tech. Rep. 1)

TAYLOR, D. W. Problem solving by groups. In *Proceedings of the XIV International Congress of Psychology, 1954.* Amsterdam: North Holland Publishing, 1954.

TAYLOR, D. W., & FAUST, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. *J. exp. Psychol.*, 1952, **44**, 360–368.

TAYLOR, D. W., & MCNEMAR, OLGA W. Problem solving and symbolic processes. In C. P. Stone (Ed.), *Annual review of psychology.* Stanford: Annual Reviews, Vol. 6, 1955.

WATSON, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, **23**, 328–336.

Received April 14, 1958.

# DIFFERENCES IN GRADIENTS OF STIMULUS GENERALIZATION AS A FUNCTION OF PSYCHIATRIC DISORDER[1, 2]

IRWIN J. KNOPF[3] AND ROBERT E. FAGER

*State University of Iowa*[4]

STIMULUS generalization (SG) is well recognized as an important explanatory concept in both infrahuman and human learning. Dollard and Miller (1950), Miller and Kraeling (1952), and Miller and Murray (1952), among others, have been instrumental in highlighting the significant role SG has come to play in personality theory, psychopathology, and psychotherapy. In the area of psychopathology, Dollard and Miller utilize the principle that "increasing the strength of the drive raises the entire gradient of generalization, increasing the strength of all generalized responses and the range of stimuli that will elicit them" (p. 178) for the purpose of explaining the development of symptoms and defense mechanisms. They also indicate that such emotional states as fear, anxiety, and guilt can acquire the properties of a drive. From this it may be postulated that increments in these emotional states raise the gradients of generalization. If one could infer from clinical observations a broad relationship between the intensity of such emotional state and severity of psychopathology, then it seems reasonable to hypothesize a relationship between severity of psychopathology and SG.

The present research investigates the hypotheses that there are differences in the shape of SG gradients among psychotic, neurotic, and control *S*s. The specific predictions are as follows:

1. Psychotics show more generalization than either neurotics or control *S*s.
2. Neurotics show less generalization than psychotics but more generalization than controls.

We were also interested in the relative performance of a brain-damaged group, but this was a side issue, and no attempt was made to account for the performance of this group on the basis of our theoretical formulation.

## METHOD

### *Subjects*

The population included 106 psychiatric *S*s (32 psychotically depressed, 24 schizophrenics, 26 neurotics, and 24 brain-damaged) and 32 controls who were nonprofessional hospital employees. None of the psychiatric *S*s was on any treatment other than psychotherapy. With the exception of the brain-damaged group, only those psychiatric *S*s who received identical diagnoses at both admission and discharge staff conferences were selected for study. Because of the importance of laboratory data and neurological consultation in arriving at a diagnosis of brain damage, the diagnosis of many of these *S*s was deferred on admission, so the discharge diagnosis was accepted as the criterion for this group. The *S* characteristics are presented in Table 1. Statistical analysis among the five groups showed significant differences ($p < .001$) in both age and education. The controls, schizophrenics, and neurotics were younger ($p < .001$) than both the depressed and brain-injured *S*s. The controls were better educated than the other four groups ($p < .001$), while the schizophrenics had more schooling than the depressives ($p < .02$) and brain-injured *S*s ($p < .01$), and the neurotics more than the brain-injured ($p < .02$). No differences in either age or education were found between the schizophrenics and neurotics or between the depressed and brain-injured; the neurotics and depressed patients had similar educational levels. With the exception of the brain-damaged group, the number of males and females in each of the other groups was approximately equal.

### *Apparatus*

The apparatus was similar to that developed by Brown, Clarke, and Stein (1958). It consisted of seven stimulus lamps mounted in a horizontal row on a black plywood panel and spaced at eight-degree intervals. The panel was curved so that all lamps were five feet from *S*, who was seated in front of a stand on which was mounted a spring-return toggle switch. When *S* moved the switch forward or backward, the words "win" or "lose" would light up in the upper or lower *left* quadrants of a small box below the central lamp. The upper and lower *right* quadrants also contained the

[1] A modification of this paper was read at the Midwestern Psychological Association meetings in Detroit, 1958.

[2] The assistance of Carl von Bickert is gratefully acknowledged.

[3] Senior author is now at the University of Texas, Southwestern Medical School, Dallas, Texas.

[4] From the Division of Psychology, Department of Psychiatry, College of Medicine.

TABLE 1
AGE, EDUCATION, AND SEX CHARACTERISTICS FOR THE FIVE GROUPS

| Group | | Mean Age | Mean Education | Sex | |
|---|---|---|---|---|---|
| | | | | M | F |
| Control | 32 | 26.03 | 15.28 | 17 | 15 |
| Schizophrenic | 24 | 26.00 | 12.43 | 12 | 12 |
| Neurotic | 26 | 27.54 | 11.96 | 14 | 12 |
| Depression | 32 | 48.59 | 10.53 | 15 | 17 |
| Brain damage | 24 | 45.21 | 10.08 | 15 | 9 |

words "win" and "lose," and were activated by *E* to give *S* information as to the actual outcome of each trial.

### *Procedure*

The *S* was told that each of the seven lamps represented a race horse. When a lamp was lighted, it indicated that this horse was running a race against other horses not represented on the stimulus panel. The seven horses on the panel were not competing against each other. Each time a light went on, it was a new race, and *S* was supposed to guess whether the horse would win or lose that race. If *S* thought the horse would win, he was to push the switch forward. If he thought it would lose, he was to pull the switch toward him. *S* was told to remember how well each horse did in order to increase the accuracy of his bets.

Each trial (or race) consisted of the presentation of a stimulus light by *E*. *S*'s response automatically terminated the stimulus, at which time *E* activated the appropriate light to indicate the actual outcome of the race. Unknown to *S*, a predetermined schedule of 80% win for the central lamp but only 20% win for the other six lamps was used. There were 210 randomly ordered races. The response measure was the frequency of win responses to each stimulus lamp.

A Lindquist Type I trend analysis (Lindquist, 1953) was used to test the difference in the shape of the gradients elicited from the five groups.

## RESULTS

The percentages of win responses to the stimulus lights for the five groups are shown in Fig. 1. For greater clarity, this figure combines the responses to peripheral lights spaced at 8, 16, and 24 degrees from the center light. It will be noted that the percentage of win responses decreases progressively as the angle separating a given light from the center increases. This result is in keeping with those of Brown, Clarke, and Stein (1958) and of Fager and Knopf (1958), who found that gradients of generalization could be elicited on this task. The over-all analysis among the five groups indicates significant differences at the .01 level in both height and slope of the gradients. In order to determine whether the schizophrenic and the psychotically depressed groups could be combined in view of their differences in age, education, and diagnosis, an analysis of the gradients for these groups was made. No differences in either height or slope of the gradients were found.

With respect to the prediction that the SG gradient obtained from a neurotic group would fall below the gradient elicited from the psychotics and above the controls, comparisons were made between the neurotics vs. the controls and the neurotics vs. the combined psychotic group. These analyses indicated no difference between the neurotics and the controls, but a significant difference at the .001 level in the slope of the gradients between the neurotics and the psychotics. To the extent that the gradient for the neurotics falls significantly below that of the psychotics, the hypothesized relationship between severity of psychopathology and SG is supported. However, this relationship is not substantiated by the null finding in gradients between neurotics and control *S*s. Perhaps the difference in

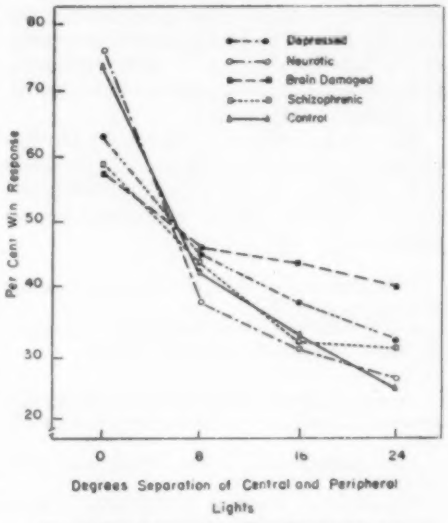


FIG. 1. PERCENTAGE OF WIN RESPONSE FOR LIGHTS AT 0, 8, 16, AND 24 DEGREES FROM CENTER LIGHT FOR PSYCHIATRIC AND CONTROL GROUPS

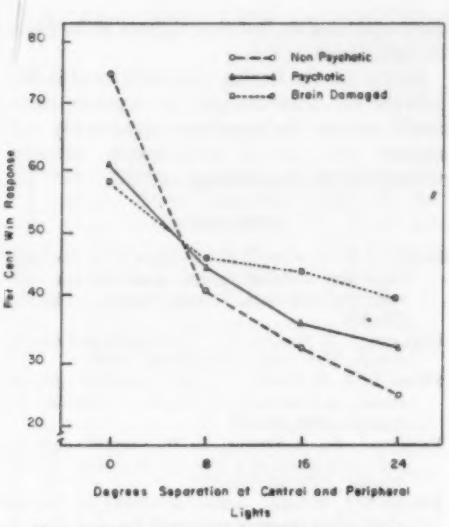


FIG. 2. PERCENTAGE OF WIN RESPONSES FOR LIGHTS AT 0, 8, 16, AND 24 DEGREES FROM CENTER LIGHT FOR PSYCHOTIC, NONPSYCHOTIC, AND BRAIN-DAMAGED GROUPS

severity of psychopathology between the neurotics and the controls was too small to be significantly manifested in generalization, or the task used to elicit gradients of SG was not sufficiently sensitive to reflect the difference.

Fig. 2 presents the percentage of win responses plotted for the brain-damaged group, the psychotic group (schizophrenic and depressed), and the nonpsychotic group (neurotic and control). The difference in the slope of the gradients between the psychotics and the nonpsychotics is significant at the .001 level. The psychotics showed a greater deviation from the predetermined schedule of 80% win to the center light and 20% win to the peripheral lights than the nonpsychotics. This finding supports the hypothesis predicting that psychotics show more generalization than nonpsychotic *Ss*. Here, again, the hypothesized relationship between severity of psychopathology and SG seems to be substantiated.

The brain-damaged group showed significantly more generalization than the control group at the .001 level. However, there was no evidence of differences between them and the psychotics. It is of interest to note that the findings for the brain-damaged *Ss* are at variance with those reported by Mednick (1955). He found that brain-damaged *Ss* showed significantly less generalization than a nonpsychiatric control and a schizophrenic group. Although the differences in sampling, experimental design, and the SG task between the two studies makes direct comparisons difficult, the apparent polarity of these findings requires definitive investigation.

## DISCUSSION

Except for the null finding between the neurotics and the controls, our results confirm the hypotheses under study and support the posited relationship between severity of psychopathology and SG. It might appear, at first blush, that age and education differences among the groups could account for a large portion of the variance in our findings. While it was not possible to control for these differences, there are some indications in the data which tend to argue against such an interpretation. For example, if age was an important contributor, then we would expect to find similar gradients among the schizophrenics, neurotics, and controls, and differences in the gradients between the schizophrenics and depressives. These expectancies are not borne out in the results. Similarly, if education was an important factor, then we would expect to find no differences in gradients between the neurotics and schizophrenics or between the neurotics and depressives, but differences in gradients between the neurotics and controls and the schizophrenics and depressives. Here, again, none of these expectations were substantiated. The data reflect no consistent relationship between gradients of SG and either age or education.

The results may plausibly be attributed to the greater amount of behavior disorganization typically observed in the more severely ill. This interpretation suggests that increments in disorganization are associated with the greater number of errors (heightened gradients of SG) obtained in the psychotics as contrasted to the neurotics and controls. Such an interpretation is consistent with our application of drive theory in that there is considerable evidence which indicates that under conditions of heightened drive and emotionality, *Ss* make more errors (become more disorganized) on

complex tasks (Farber & Spence, 1953; Knopf, Worell, & Wolff, 1958; Kohn, 1954; Korchin, 1955; Taylor & Spence, 1952). Thus, the concept of drive level may have some utility in further understanding the behavior of psychopathological groups. However, the ultimate utility of this concept rests upon further extensions of the relationship between degree of psychopathology and a wide variety of performances from which drive level may be inferred.

The performance of the brain-damaged *Ss* is quite interesting in that these individuals show elevated gradients comparable to those of the psychotics. This finding suggests that the brain-damaged group, at least with respect to this SG task, is as disorganized as the psychotics. While it seems reasonable to attribute this disorganization to neurological deficits, Miller (1958) recently referred to a growing body of literature which indicates that brain lesions can modify emotionality, and brain stimulation can produce states from which drives can be inferred. Thus, future research may demonstrate that the disorganized behavior of the brain-damaged group may be in part a function of changes in drive level associated with neurophysiological impairment.

## SUMMARY

On the basis of a postulated relationship between severity of psychopathology and SG, it was hypothesized that psychotics would show more generalization than nonpsychotic *Ss* and neurotics would show an SG gradient falling below that of the psychotics and above that of the controls. An SG task was administered to a total of 138 *Ss* (32 nonprofessional hospital employees, 26 neurotics, 32 psychotically depressed, 24 schizophrenics, and 24 brain-damaged). The results indicate that (*a*) the psychotics showed more generalization than the nonpsychotics; (*b*) the neurotics showed less generalization than the psychotics, but did not differ from the controls; and (*c*) the brain-damaged *Ss* showed more generalization than the nonpsychotics, but they did not differ from the psychotics.

Except for the finding of no difference in SG between the neurotics and the controls, these results confirm the hypotheses under study and support the posited relationship between severity of psychopathology and SG.

## REFERENCES

BROWN, J. S., CLARKE, F. C., & STEIN, L. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55**, 359–362.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

FAGER, R. E., & KNOPF, I. J. Relationship of manifest anxiety to stimulus generalization. *J. abnorm. soc. Psychol.*, 1958, **57**, 125–126.

FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.

KNOPF, I. J., WORELL, JUDITH, & WOLFF, H. D. The effect of an "anxiety reducing" drug on stimulus generalization under experimental stress. *Amer. Psychologist*, 1958, **13**, 343. (Abstract)

KOHN, H. The effect of variations of intensity of experimentally induced stress upon certain aspects of perception and performance. *J. genet. Psychol.*, 1954, **85**, 289–304.

KORCHIN, S. Perceptual adequacy in life stress. *J. Psychol.*, 1955, **39**, 109–116.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

MEDNICK, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51**, 536–542.

MILLER, N. E. Central stimulation and other new approaches to motivation and reward. *Amer. Psychologist*, 1958, **13**, 100–108.

MILLER, N. E., & KRAELING, DORIS. Displacement: Greater generalization of approach than avoidance in a generalized approach-avoidance conflict. *J. exp. Psychol.*, 1952, **43**, 217–221.

MILLER, N. E., & MURRAY, E. J. Displacement and conflict: Learnable drive as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.*, 1952, **43**, 227–231.

TAYLOR, JANET A., & SPENCE, K. W. Relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

Received April 23, 1958.

# ATTRACTIVENESS OF GROUP AS FUNCTION OF SELF-ESTEEM AND ACCEPTANCE BY GROUP[1]

JAMES E. DITTES
*Yale University*

THE power of a group to exert influence over a member depends largely on the degree to which the person values his membership in the group (Back, 1951; Dittes, 1959; Dittes & Kelley, 1956; Festinger, 1950; Festinger, Schacter, & Back, 1950; Kelley & Shapiro, 1954). Two determinants of this variable—known variously as valuation of group membership, attraction toward a group, or cohesiveness—are studied here.

A person's attraction toward membership in a group, like motivational attraction toward any object, may be considered a function of two interacting determinants: (*a*) the extent to which his particular needs are satisfied by the group, and (*b*) the strength of his needs (Cartwright & Zander, 1953).

Assuming a multiplicative relation between need and gratification, several predictions may be derived from this formulation. These indicate that the effectiveness of either of the variables—strength of need or level of available gratification—depends on the level of the other. The relation of attraction with either variable may be direct, zero, or inverse, according to whether the level of the other variable is positive, zero, or negative. The specific predictions of interaction are stated below and illustrated in Fig. 1.

1. If a person has a relatively strong need, his attraction to the group is likely to vary directly with the extent to which the group satisfies that need.

2. If an individual's need is low, the degree of gratification afforded by the group is a relatively less important determinant of attraction. In the extreme case of a person with a need at a zero level, the amount of potential gratification or frustration of that need by the group has no effect on his level of attraction toward the group.

3. If a group offers relatively great satisfaction of a particular need, a person's attraction to the group should vary directly with the strength of that need.

4. If a group definitely frustrates a need, attraction should vary inversely with the strength of the need. A person with a stronger need should experience greater punishment in the frustration and consequently reduce his attraction toward the group by the greatest amount. If attraction in any instance were dependent solely on this single need, its frustration would drive an individual to leave the group. In most instances, attraction depends on many other factors and may remain positive, even though sharply reduced by the frustration of a strong need. The zero level of attraction in Fig. 1, around which the variation due to a single need is plotted, actually may be expected to be a positive value as a result of the effect of other factors.

Perhaps the type of gratification that persons most commonly gain from membership in groups is that of social acceptance and the closely related rewards of support, recognition, security, and general esteem from other members of the group. The personal need that is gratified by acceptance may be variously designated and measured. It is assumed here that the need is best indicated by a person's level of self-esteem or general sense of adequacy. The lower the level of self-esteem, the greater is a person's need for such supports to self-esteem as are provided by acceptance in a group.

Earlier results (Dittes & Kelley, 1956; Festinger et al., 1950; Jackson, 1952; Kelley, 1951; Kelley & Shapiro, 1954; Zanders & Cohen, 1955) clearly indicated that attraction toward the group functions directly with the level of acceptance experienced. This relationship is consistent with the first prediction above, assuming that subjects had relatively strong needs for acceptance. However, the interactions proposed in the analysis just presented have not been investigated. Three hypotheses concerning the interaction of level

[1] Appreciation is due Harold H. Kelley for suggestions on the design of the experiment and on the manuscript.

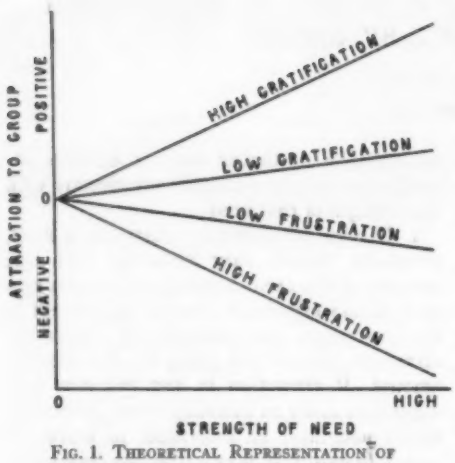


FIG. 1. THEORETICAL REPRESENTATION OF ATTRACTION TO A GROUP AS A FUNCTION OF THE LEVELS OF A PARTICULAR NEED OF AN INDIVIDUAL AND OF THE GRATIFICATION OF THE NEED BY THE GROUP

of acceptance with the personality characteristic of self-esteem may be developed from the second, third, and fourth general deductions:

*Hypothesis I.* The tendency of greater acceptance in a group to produce greater attraction toward the group is greater among persons with low self-esteem than among persons with high self-esteem.

*Hypothesis II.* Among persons experiencing acceptance in a group, attraction to the group is greater among persons with low self-esteem than among persons with high self-esteem.

*Hypothesis III.* Among persons experiencing nonacceptance in a group, attraction to the group is less among persons with low self-esteem than among persons with high self-esteem.

## METHOD

*Subjects*, male college freshmen, met in small groups of five or six members each for a two-hour discussion and task session. The activities of the groups are more fully described elsewhere (Dittes, 1959). In all, 104 *Ss* met in 20 different groups. The experimental procedures were replicated in each group. Members of each group were not previously well acquainted with each other.

So that attractiveness of the group could be seen varying around a relatively high positive level, several motivations were introduced to make the group highly attractive. These included the use of inherently interesting topics and tasks, assignment of a group goal offering a money reward and prestigeful recognition for successful attainment, and an atmosphere of prestige concerning the experiment as a whole, fostered by such devices as making a tape recording of the group's discussion, as though to be studied later by social scientists. Without this added motivation, if attractiveness of the group had depended solely on the acceptance variable, the *Ss* for whom this variable was expected to have a negative effect might have left the group, either physically or by some psychological escape. An artificial floor might have been introduced at the zero level of attraction, preventing a measure of variation below this level.

*Different levels of apparent acceptance* were introduced by letting *Ss* see ratings of themselves which they believed had been made by other members of the group. During the first hour, the group's discussion was interrupted three times to allow members to rate the desirability of having each other member in the group. These ratings were requested by *E* as though for his own interest. At an "intermission," the *Ss*' interest in these ratings was aroused, and each person was allowed to see privately what he believed to be the ratings of desirability made of him by other members of the group. Actually, the distributed ratings were fictitious and had been prepared in advance to induce, at random, different levels of feeling accepted.

It was attempted to introduce three levels of acceptance in each group, represented by ratings clustered, respectively, well above the average, at the average, and well below average. However, subsequent questionnaire items measuring the *Ss*' direct reactions to these ratings indicated that the intended high and average conditions evoked very similar reactions (Dittes, 1959). Both these groups of *Ss* felt pleased, accepted, and generally secure. Accordingly, the data from both these groups are combined and reported together as the *satisfying* condition. *Ss* receiving the ratings below average reported feeling disturbed, nonaccepted, and threatened. This group will be known as the *frustrating* condition.

After the *Ss* saw the ratings, they continued to function as a group, but without any further actual interaction, which might have diluted the feelings of acceptance or nonacceptance induced by the ratings. *Ss* continued in a face-to-face group around a table but performed written tasks individually, with the expectation that some of this written work would provide the basis for further discussion. Actually, the final discussion consisted, instead, of a disclosure of the purposes and procedures of the experiment.

*Self-esteem* was measured with an index made up of three components. (*a*) A questionnaire completed before the experimental session included 11 items asking *Ss* to indicate on a six-point scale whether particular statements describing feelings of adequacy and self-esteem applied to them more or less than to the average man their age. The odd-even reliability of these items, corrected for length, yielded a product-moment correlation of .82. (*b*) On the questionnaire at the end of the session, before the disclosure of the procedures, six items asked *Ss* about their general sense of adequacy among groups of peers. Replies to these differed sig-

nificantly between the two experimental conditions (providing one index of the valid effectiveness of the experimental manipulation). These differences were removed by taking as each *S*'s score his deviation from the mean of his experimental condition. These scores were significantly correlated with the presession measure of self-esteem. (*c*) It was reasoned that the actual ratings *S*s received from each other (but did not see) might provide an indication of self-esteem, since they were based on brief acquaintance and would perhaps depend chiefly on the kind of behavior (e.g., active participation vs. reticence, or steady, relevant contribution vs. defensive effusiveness) which is largely a function of self-esteem. These ratings were, in fact, found to correlate significantly with the self-ratings and were accordingly included in the index. The scores on the three components were summed to provide the general index of self-esteem.

For purposes of comparing different levels of self-esteem, the distribution of scores on the index was examined and found to fall into three fairly distinct clusters. The scores were accordingly divided into equal thirds, representing high, average, and low levels of self-esteem. This division was made before the results here reported were examined.

*Attraction to the group* was measured with three items on the final questionnaire. The items were as follows:

> If this group should be invited to meet again or to serve as a demonstration panel, would you want to continue working with it, or would you just as soon not attend any further meetings with this group?
>
> How much do you like participating in this experiment?
>
> If this group is invited to meet again, but after the experimenter examines the records of this session, you are not invited to participate, how disappointed will you be?

*S*s indicated replies on seven-point graphic rating scales. Scores on the three scales were significantly intercorrelated and were summed to provide the measure of attraction to the group. The 20 experimental groups differed substantially, although not significantly, in their average level of attraction. Presumably this was a result of variation from session to session in the effectiveness of the incentives intended to create a generally high level of attractiveness. To correct for this variation, the deviation of each *S* from the mean of his group was taken as his score.

## RESULTS AND DISCUSSION

*S*s made to feel accepted—the satisfying condition—reported being significantly more attracted to the group than *S*s made to feel nonaccepted—the frustrating condition ($t = 2.98$, $p < .01$).[2] For the three intended conditions of high, average, and low acceptance, the mean scores were .7, .6, and −1.4. Scores for the intended high and average conditions differed slightly in the predicted direction but, as with the previously reported results of direct effects, were affected in substantially the same way by the experimental manipulations. They are therefore appropriately considered together as a single condition. The difference in attraction between accepted and nonaccepted persons repeats previous findings (Dittes & Kelley, 1956; Jackson, 1952; Kelley & Shapiro, 1954).

However, as predicted by Hypothesis I and as illustrated in Fig. 2, this difference was primarily due to the *S*s with lowest self-esteem, presumably indicating strongest needs for social acceptance. The $p$ values for the difference between the satisfying and frustrating conditions at the three successive levels of self-esteem were $> .50$, $< .20$, and $< .002$, respectively. The experimental conditions of high and low acceptance produced a difference among persons with low self-esteem that was significantly greater than the difference among persons with high self-esteem; the difference of the differences yielded a $t$ of 2.02, $p < .05$.

Hypothesis II predicted that for persons feeling accepted, attraction should be greater among those with greater need for acceptance, as indicated by lower self-esteem. The results, shown by the upper curve in Fig. 2, were in the predicted direction but did not approach statistical significance. The difference between

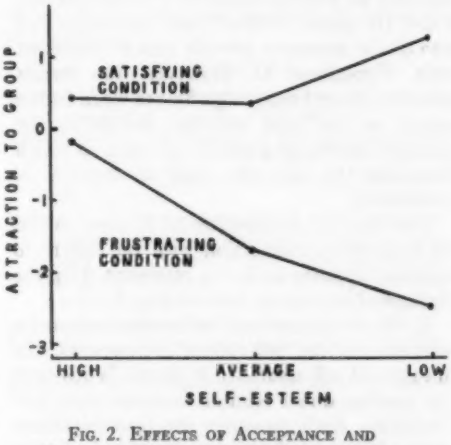


FIG. 2. EFFECTS OF ACCEPTANCE AND NONACCEPTANCE ON ATTRACTION TO GROUP AMONG PERSONS WITH DIFFERENT LEVELS OF INITIAL SELF-ESTEEM

[2] All probability levels are based on two-tail tests of significance.

TABLE 1
ATTRACTION SCORES FOR SUBJECTS HIGH AND LOW ON EACH COMPONENT MEASURE OF SELF-ESTEEM

| Experimental Condition | Self-ratings | | | | Others' Ratings of Esteem | |
|---|---|---|---|---|---|---|
| | General Self-esteem | | Self-confidence in Groups | | | |
| | High | Low | High | Low | High | Low |
| Satisfying | .4 | .9 | 1.0 | .2 | .2 | 1.1 |
| Frustrating | −1.2 | −1.6 | −.2 | −2.7 | −1.1 | −1.6 |
| Difference between conditions | 1.6 | 2.5 | 1.2 | 2.9 | 1.3 | 2.7 |

the high and low levels of self-esteem produced a $t$ of 1.07, $p < .30$.

Hypothesis III was given clearer, though still tentative support. In the frustrating condition of nonacceptance, attraction to the group was directly related to level of self-esteem. *S*s with lowest self-esteem reported the least attraction. The difference between the high and low levels of self-esteem produced a $t$ of 1.8, $p < .10$.

It may be asked whether the results are attributable largely to the effect of any one of the three components of the self-esteem index. Table 1 presents the attraction scores for *S*s above and below the median on each of the three measures that comprised the index of self-esteem. Trends consistent with Hypotheses I and III appear with all three measures, and two of the measures provide trends consistent with Hypothesis II. However, the trends generally do not reach significance when *S*s are sorted on any one measure, indicating the greater stability of a combined index, in which error due to any one type of measure is minimized.

Two facts of methodological interest should be reported in connection with the failure to confirm Hypothesis II—as represented by the flatness of the upper curve in Fig. 2.

1. The predicted relation between attraction and one of the self-esteem components, the measure of self-assurance in groups, is obscured by another direct relation between these two variables. Both measures are probably functions, in part, of acceptance experienced in previous groups. Persons used to being accepted in groups should report more assurance in groups and also like being in groups. Therefore some correlation is to be expected between these two measures, and a slight, though significant correlation was found ($r = .20$, $p < .05$) even when differences due to experimental manipulations were removed for each measure by taking scores as deviations from the means of each experimental condition. As can be seen in the middle section of Table 1, this correlation tended to enhance the experimental effect within the frustrating condition, making it highly significant ($p < .01$ for the difference between those high and low in self-assurance in groups). But the correlation reversed the predicted effect in the satisfying condition (Hypothesis II).

This correlation is not to be expected and was not found with the two other more general estimates of self-esteem, because they depend on support and other rewards in a wider range of experiences than small groups. When an index composed of just these two measures is used to divide *S*s at the median into those with high and low self-esteem, as in Table 2, then those in the satisfying condition show a significant trend in the expected direction.

2. Efforts to avoid a floor effect in attractiveness, by offering various incentives to increase the general attractiveness of the group, may actually have produced a ceiling effect. There is likely to be a maximum attractiveness possible for such an experimentally created group, a maximum imposed largely by students' other interests and obligations. Regardless of the attractiveness of the group, other commitments impose a fixed upper limit on an *S*'s willingness to express a desire to meet again with the group. The various incentives employed may have raised general attractiveness near this level, so that it was difficult to register addi-

TABLE 2
ATTRACTION SCORES FOR SUBJECTS HIGH AND LOW ON TWO-COMPONENT INDEX OF SELF-ESTEEM

| Experimental Condition | Self-esteem | | $P$ Value |
|---|---|---|---|
| | High | Low | |
| Satisfying | −.2 | 1.5 | <.05 |
| Frustrating | −1.0 | −1.8 | n.s. |
| Difference between Conditions | .8 | 3.3 | .05 |

TABLE 3
ATTRACTION SCORES IN SATISFYING CONDITION IN GROUPS WITH HIGH AND LOW AVERAGE ATTRACTION

| | Self-esteem (3-component index) | | | | | Self-esteem (2-component index) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | High | Average | Low | Low vs. High and Ave. | | High | Low | Low vs. High | |
| | | | | Diff. | *P* | | | Diff. | *P* |
| In groups with above-average attraction scores | .8 | 1.4 | .2 | −.9 | n.s. | .5 | 1.0 | .5 | n.s. |
| In groups with below-average attraction scores | 0 | −2.2 | 3.6 | 5.1 | <.02 | −2.5 | 2.5 | 5.0 | <.02 |
| Difference between groups | | | | 6.0 | <.01 | | | 4.5 | .02 |

tional upward variation in attraction due to satisfaction of high need for acceptance.

This possibility can be checked by comparing data from experimental groups averaging high attraction and those averaging low attraction. Fourteen of the groups had average attraction scores compactly arrayed within two points above the over-all mean, suggesting the existence of a kind of ceiling. The remaining six groups ranged in average attraction from one to six points below the mean. As Table 3 shows, the trends in the satisfying condition for attraction to increase with lower self-esteem are due almost entirely to the six groups with below-average attraction. In these groups, with either index of self-esteem, persons with low self-esteem were significantly more attracted to the group than those with high self-esteem, when they were in the satisfying condition. This difference is significantly greater than the corresponding difference in groups above the mean on attraction.

It also must be conceded that the type of experimental manipulation used, based exclusively on written ratings, may be unable to effect the most powerful levels either of gratification or of frustration. The subtle cues of word, gesture, and expression on which persons most commonly rely as signs of acceptance were not involved here. This kind of written communication may at best produce conditions corresponding more to the low levels of gratification and frustration represented in Fig. 1. Hence, the resulting curves would necessarily be flatter than if the full range of powerful rewards and frustrations usual in interpersonal interactions were invoked.


## SUMMARY

Persons made to feel well accepted in a group found the group more attractive than did those made to feel poorly accepted. But this difference was significantly greater among persons with low self-esteem than among persons with high self-esteem, low self-esteem being taken as an indication of strong need for acceptance. Some support was given to the additional predictions that attractiveness of membership in a group varies directly with need for acceptance when a group is accepting and inversely when a group is nonaccepting. The results were interpreted as illustrating a multiplicative relation between two factors—strength of need in the individual and level of gratification available in the group—as determinants of the attractiveness of a group and, hence, of the power of the group to influence the individual.


## REFERENCES


BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

CARTWRIGHT, D., & ZANDER, A. Group cohesiveness: Introduction. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. Evanston: Row, Peterson, 1953. Pp. 73–91.

DITTES, J. E. Effects of changes in self-esteem upon impulsiveness and deliberation in making judgments. *J. abnorm. soc. Psychol.*, 1959, **58**, 348–356.

DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J abnorm. soc. Psychol.*, 1956, **53**, 100–107.

FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.

FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.

JACKSON, J. Analysis of interpersonal relations in a formal organization. Unpublished doctoral dissertation, Univer. of Michigan, 1952.

KELLEY, H. H. Communication in experimentally created hierarchies. *Hum. Relat.*, 1951, **4**, 39–56.

KELLEY, H. H., & SHAPIRO, M. M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *Amer. sociol. Rev.*, 1954, **19**, 667–677.

ZANDER, A., & COHEN, A. R. Attributed social power and group acceptance: A classroom experimental demonstration. *J. abnorm. soc. Psychol.*, 1955, **51**, 490–492.

Received April 23, 1958.

# PROCESS-REACTIVE SCHIZOPHRENIA AND GENETIC LEVELS OF PERCEPTION

HAROLD J. FINE

*Veterans Administration, Bridgeport, Connecticut*

AND CARL N. ZIMET

*Yale University*

THE terms *process* and *reactive* schizophrenia have been widely used to differentiate two parameters of dementia praecox. Process schizophrenia is thought to be a disorder of gradual or insidious development in which prodromal signs of the illness are present early in life. Typically, there is an absence of precipitating factors, a failure to interact with peers and excessive fantasy in childhood, and a dull, rigid, or inappropriate affect apparent relatively early in the lifetime of these individuals. Withdrawal, autistic logic, and other aspects of a deep narcissistic regression are noted in social service histories. Economic, social, and sexual adjustment is inchoate and in many ways as primitive as the thinking disorder. When and if the acute "break" occurs, the patient tends to deteriorate progressively, sometimes in spite of all attempts at treatment. Eventually, a good proportion of these patients form the population of "burned out" schizophrenics with little likelihood of recovery.

The reactive schizophrenics, on the other hand, show a relative absence of the prodromal signs noted in the process group. In general, there is an adequate social, economic, and sexual adjustment. A noteworthy precipitating stress can characteristically be identified, and over-all prognosis for recovery is good. The contrast between these two groups, it must be noted, is not simply a difference in severity of illness; the reactive schizophrenic may often be more bizarre in his behavior during the acute stage of his psychosis.

This hypothesis of two kinds of schizophrenia suggests the possibility of different etiologies or, as Becker (1956) has pointed out, of differences in "end points on a continuum of levels of personality organization." A possible organic basis for process schizophrenia is also tenable. Brackbill and Fine (1956) found, for example, that the process schizophrenic reacts to a perceptual task like the Rorschach in a manner more similar to that of known central nervous system patients than do reactive schizophrenics.

In an attempt to evaluate further the process and reactive schizophrenias, a study of the level of perceptual organization was conceived in terms of Werner's (1948) ontogenetic principle that development occurs from a state of grossness and lack of differentiation to a state of increasing differentiation and hierarchic integration. The undifferentiated structure or process is said to be syncretic, diffuse, labile, indefinite, and rigid. When differentiated and organized, the terms discrete, articulate, stable, definite, and flexible are applicable. Recently, Friedman (1953), Hemmendinger (1953), Siegal (1953), and others attempted to link some of these ontogenetic principles with clinical theory by applying Werner's developmental approach to the structural aspects of Rorschach scores. A scoring system for Rorschach response location was developed to reflect genetic changes. The system includes differentiation, defined in terms of the amount of separation of the blot areas into different parts, their hierarchic integration, and the organization of these areas into a combined whole. Such a scoring system enables responses to be distinguished into "genetically low" and "genetically high" scores. These investigators found that more seriously impaired clinical groups showed a preponderance of genetically lower responses, and as one went up the scale to the less severely impaired clinical groups, the frequency of higher genetic responses increased.

From these considerations and observations, the following hypotheses were formulated: (*a*) Patients identified as process schizophrenics show more indices of perceptual immaturity as defined by Rorschach criteria than those identified as reactive schizophrenics. Conversely, (*b*) reactive schizophrenics reflect

more perceptual maturity as defined by Rorschach developmental scores.

## METHOD

The two groups used here comprised 36 patients diagnosed as process schizophrenics and 24 as reactive schizophrenics. The patients in these groups were all those World War II veterans admitted to a VA neuropsychiatric hospital during the period of a year who were diagnosed as schizophrenic and who were given a battery of psychological tests including the Rorschach. Cases involving multiple diagnoses, latent schizophrenia, or borderline states were eliminated. The selection procedure was that described by Kantor, Wallner, and Winder (1954).

Patients were classified as reactive or process on the basis of case history information alone without knowledge of the Roschach or regard for the severity of the clinical symptomatology. Case history data were drawn from four areas of development covered in social service reports: birth to fifth year, fifth year to adolescence, adolescence to adulthood, and adulthood. Items in these areas revolved around the intrapsychic, social, sexual, and familial spheres. Samples were reliably differentiated into process and reactive categories ($p$ = .01) by Kantor, Wallner, and Winder (1954). For the present study, only those *S*s were included about whom there was complete agreement among the judges as to the type of schizophrenia (60 out of 86 patients).

The method of analyzing the Rorschach protocols followed the system developed by Friedman (1953), which reflects levels of perceptual development as suggested by Werner (1948). In evaluating the reliability of this scoring system, Friedman found a high degree of agreement among four judges. Out of 1136 responses, agreements ranged from 89.7% to 95.5%. The following scores were used: *DW*, confabulatory response; *minus* a response with inadequate form quality; *amorphous*, a response in which the shape of the blot plays no determining role; *vague*, a response involving a diffuse general impression of the blot; *mediocre*, a response in which the gross outline is taken into account so that specific form implied in the content matches the blot; *plus*, a response in which two or more discrete areas are combined into one percept; *plus-plus*, a response in which a unitary area or the whole blot is perceptually articulated and reintegrated into a well differentiated, unified percept.

The medians for the seven score categories presented above were computed for the combined sample, and chi square analyses were run on the basis of frequencies above and below the median in the process and reactive groups. In addition, each of the responses was designated as immature or mature, based on combining *DW*, *amorphous*, *vague*, and *minus* into the immature category and *mediocre*, *plus*, and *plus-plus* into the mature category. For the analysis of differences between process and reactive groups, the $t$ test was used. The Rorschach protocols were coded in order to avoid identification of the process and reactive respondents. Each of the authors scored the protocols independently; the percentage of agreeement for the seven score categories was 92.7.

There was no evidence that the two groups of patients differed in Wechsler IQs. Similarly, no significant differences were found in the number of Rorschach responses in the two samples.

## RESULTS

The results indicate that the two groups react to a perceptual task like the Rorschach in the predicted direction. When all of the scores of the process schizophrenics were dichotomized into mature and immature combinations and all the scores of the reactive schizophrenics were dichotomized into mature and immature combinations as specified above, $t$ for the mean immature responses was 3.07, significant beyond the .01 level. Fifty-seven per cent of the responses of the process group were in the immature category, whereas 20% of the responses of the reactive schizophrenics were in this same category.

An analysis of individual score categories reveals the following: The confabulated responses *DW* show no significant difference between the two groups ($p$ = .60). In terms of absolute frequencies, however, the process group produced seven *DW*s, whereas the the reactive group produced two. The *minus* and the *amorphous* categories both show significant differences between the two groups beyond the .01 level. No significant differences were found in the *vague* category ($p$ = .70) or in the *mediocre* class of responses ($p$ = .10). The *plus* category was significantly different beyond the .02 level, the reactives responding with more of these integrated perceptions. The

TABLE 1

DIFFERENCES IN RORSCHACH DEVELOPMENT SCORES FOR PROCESS AND REACTIVE SCHIZOPHRENICS

| | $\bar{x}$ | $SD$ | $t$ | $p$ |
|---|---|---|---|---|
| Mature–Immature | | | 3.068 | .001 |
| Process | .569 | 1.860 | | |
| Reactive | .201 | .151 | | |

| Subscore Category | $\chi^2$* | $p$ |
|---|---|---|
| *Plus-Plus* | .904 | .30 |
| *Plus* | 4.230 | .02 |
| *Mediocre* | 1.740 | .10 |
| *Vague* | .003 | .70 |
| *Amorphous* | 6.163 | .005 |
| *Minus* | 9.116 | .001 |
| *DW* | .294 | .60 |

* All chi square values are based on $df$ of 1 and $N$ of 60.

chi square for the *plus-plus* category was not significant, but the occurrence of these scores was too infrequent to permit an adequate test.

## DISCUSSION

The data indicate greater perceptual immaturity among the process schizophrenics and more adequate and integrated perceptual functioning in the reactive group.

In view of the fact that the genetic scoring categories reflect successively higher levels of integrated functioning, ranging from autistic, syncretic, and physiognomic to organized, realistic, and differentiated, one might use the conceptualization of primary and secondary process thinking. In other words, the perceptions scored as immature and mature might also be regarded respectively as expressing primary and secondary levels of thinking. Primitive thought processes appear to break through more freely in process schizophrenia. Of note here is the *minus* category, which in some sense may mirror defective reality testing by a reversion to a more primitive model of thinking both genetically and structurally. As Rapaport (1954, pp. 271–272) has stated, "Structurally, the emergence and development of thought processes is one aspect of the emergence and development of the ego and thought processes are integral parts of ego processes. They are an outstanding aspect and instrument of reality testing, subserving the reality principle." One interpretation of the data may thus be based on a translation of these genetic scores into the realm of ego psychology as it reflects ego development.

Another difference noted qualitatively was that the process schizophrenic tends to be more concrete and more stimulus bound, while the reactive seems more capable of abstraction. There is reason to believe that one of the critical differences between the process and reactive groups is in terms of a kind of thinking disorder.

The conceptualization of the process-reactive classification of schizophrenia which has been supported here on the basis of psychological test findings has also received support through findings of differences in physiological responsiveness (King, 1958). The data do not suggest etiological formulations (e.g., constitutional versus psychogenic), but questions must be raised as to whether the process-reactive classification can serve to provide some clues about etiology. What has been identified in this paper are two possible species of schizophrenia. This is not to say that these are the only two species, nor do they seem to be, in view of the degree of overlap in the findings, mutually exclusive categories. While the results of this study are significant, the predictive efficiency of the individual genetic scores is rather low. This should be a point to bear in mind in erecting diagnostic or prognostic indicators based on this classification in a clinical setting. A more thorough investigation of etiology via the intensive longitudinal approach is called for to help sort out constitutional factors, psychogenic factors, or the interaction between these two that might contribute to the development of schizophrenia.

## SUMMARY

An attempt was made to evaluate the degree of perceptual regression, as defined by Rorschach responses, in reactive and process schizophrenics. The hypothesis that process schizophrenics show a preponderance of genetically lower responses than reactive schizophrenics was confirmed. The data indicate grosser perceptual immaturity within the process schizophrenic group and more adequate and integrated perceptual functioning within the reactive group. Several possible interpretations of these findings are presented.

## REFERENCES

BECKER, W. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.

BRACKBILL, G., & FINE, H. J. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, **52**, 310–313.

FRIEDMAN, H. Perceptual regression in schizophrenia, a hypothesis suggested by the use of the Rorschach test. *J. proj. Tech.*, 1953, **17**, 171–185.

HEMMENDINGER, L. Perceptual organization and development as reflected in the structure of the Rorschach test responses. *J. proj. Tech.*, 1953, **17**, 162–170.

KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1954, **17**, 157–162.

KING, G. F. Differential autonomic responsiveness in

the process-reactive classification. *J. abnorm. soc. Psychol.*, 1958, **56**, 160–164.

RAPAPORT, D. On the psychoanalytic theory of thinking. In R. P. Knight (Ed.), *Psychoanalytic psychiatry and psychology*. Vol. 1. New York: International Universities Press, 1954. Pp. 259–273.

SIEGAL, E. L. Genetic parallels of perceptual structurization in paranoid schizophrenia: An analysis by means of the Rorschach technique. *J. proj. Tech.*, 1953, **17**, 151–161.

WERNER, H. *Comparative psychology of mental development*. Chicago: Follet, 1948.

Received May 1, 1958.

# A METHOD FOR DETERMINING TYPES OF SELF-ESTEEM[1]

STANLEY COOPERSMITH

*Wesleyan University*

ALTHOUGH self-esteem is generally assumed to be a major factor in determining behavior, there has been relatively little research directed towards clarifying its significance and dynamics. Studies in such diverse areas as recall of completed and incompleted tasks (Rosenzweig, 1938), level of aspiration (Sears, 1940), reactions to threat (Lazarus & Longo, 1953), and responses to conformity pressures (Asch, 1948) have concluded that self-esteem is a significant contributing variable. In the absence of an adequate method of determining self-esteem, however, there have been relatively few direct studies of this variable. The present study was conducted to develop measures capable of distinguishing between subjects high and low in self-esteem, and between subjects exhibiting reality-based and defensive responses.

Studies of self-esteem are difficult to evaluate, since the term itself is vague and subject to manifold interpretations. Underlying the variety of terminology—pride (Baldwin & Levin, 1957), ego (Freud, 1927; Gough, 1954), dominance (Gough, 1954; Maslow, 1939), self-assertion (Gough, 1954), self-cathexis (Jourard, 1957)—is a distinction between self-evaluation and the manner in which this self-evaluation is expressed in behavior. Various definitions emphasize short- and long-term, hierarchical, ego, behavioral, social, and self states, all of which presumably have in common the ego needs postulated by Freud (1927) and extended and clarified by A. Freud (1938), Horney (1937) and other neo-Freudians, and Rogers (1942). In the present experiment, self-esteem is defined from both the perspective of the subject, via the Self-Esteem Inventory (SEI), and from that of the observer, via the rating of his self-esteem behavior, on the Behavior Rating Form (BRF). Types of self-esteem are then derived from various combinations of SEI and BRF scores. Two of the major perspectives of self-esteem, the subjective and the behavioral, are measured in this study, while a third, the unconscious aspect, is generally acknowledged to be correlated with behavior.

The distinction between genuine and defensive self-esteem is based upon the extent of the differences between the subjective and behavioral evaluations. In those cases where the discrepancy is wide, the person's experiences are apparently reflected in his behavior, while his defenses distort the self-descriptions so that they bear little resemblance to his experiential background. Correspondence between the various perspectives signifies experientially based behaviors and perceptions and is generally acknowledged to be indicative of psychological health.

Since the frequency and intensity of reality experiences intrude upon the individual to such a degree that they cannot be avoided in self-evaluation, it may be assumed that although defensive and distorting factors may be present, the subjective evaluation of self-esteem is, in the majority of cases, in substantial agreement with its behavioral expression. In addition, cases of fantasy life or incorporated standards of evaluation relatively disparate from experience, and capable of remaining firm in the face of long-term contradictory evidence are assumed to be relatively infrequent. This reasoning leads to the hypothesis that persons whose experiences have been preponderantly successful should generally tend to express confidence and assurance in both their behavior and perceptions, while those who have had more failure experiences should generally tend to express either caution and hesitancy, or

[1] This paper is adapted from a dissertation submitted to Cornell University in partial fulfillment of the requirements for the doctor of philosophy degree. The study was originated under U.S.P.H.S. Grant M(901) and is currently being supported by a grant from the Connecticut Association of Mental Health.

The author wishes to express his appreciation to Alfred L. Baldwin, Robert H. Dalton, Harry Levin, and Robert B. MacLeod of Cornell University, and to the personnel and the children of the Ithaca school system for their interest and assistance.

This is the first of a series of related papers on "Self-Esteem as a Determinant of Behavior."

attention-seeking and aggression in both their behaviors and perceptions. The final hypothesis is that in cases in which there is disagreement between self-evaluation and behavioral expression there should be either (*a*) low self-evaluation due to high standards to which the individual rigidly adheres, or to failures in areas of experience not tapped in the present inventory; or (*b*) high self-evaluation due to defenses against a low self-evaluation, to successes achieved in fantasy, or to success in areas of experience not tapped in the present inventory.These hypotheses have the implication that cases in which the SEI score is high and the BRF score is low may be expected to show different profiles on a number of related motivational, ability, and experiential variables from those cases in which the SEI score is low and the BRF score is high. Similarly, different profiles may be expected between cases where both the SEI and BRF scores are high and those cases where both these scores are low.

## METHOD

The experimental procedure involved three stages: (*a*) the development and administration of a Self-Esteem Inventory and a Self-Esteem Behavior Rating Form; (*b*) the selection of samples of agreement and disagreement between self- and behavioral evaluations; and (*c*) determining whether these samples displayed significantly different patterns on objective measures of personality and academic and social success and failure.

### *Subjects*

The *S*s were 102 fifth- and sixth-grade children, 10 to 12 years of age, attending the public schools in a small Eastern city. The group consisted of 49 girls and 53 boys who resided in or near a middle-middle- to upper-middle-class neighborhood. The area is judged to be relatively stable, with the majority of its residents either associated with the local university, professionals, or proprietors of local businesses.

The age group 10–12 was chosen since the personality has been relatively well formed by this time, and the adolescent turmoil frequently noted in our society is not yet likely to have occurred. A school situation was chosen as the setting for the experiment because: (*a*) the strong cultural pressures toward academic achievement make it likely that success and failure in school would be reflected in self-esteem measures; (*b*) the body of evidence and theory developed by McClelland Atkinson, Clark, and Lowell (1953) provides a framework for studies in the area of achievement; and (*c*) the school environment provides a wide range of information regarding the abilities, motivation, and performance of the child, as well as raters (peers, teacher, principal) who have had opportunities to observe the child over long periods of time and under varied conditions.

### *Measures of Self-Esteem*

*Self-Esteem Inventory.*[1] A Self-Esteem Inventory was constructed on the basis of items selected from the Rogers and Dymond scale (1954) which were reworded for use with children. Several additional items were designed by the experimenter. The entire battery was presented to five psychologists who sorted the items into two groups—those indicative of high self-esteem and those indicative of low self-esteem. Items that appeared to be ambiguous, repetitious, or about which there was disagreement were eliminated. The set of items was then tested for comprehensibility with a group of 30 children. The final Inventory consisted of 50 items concerned with the *S*'s perceptions in four areas: peers, parents, school, and self. Each item checked in the positive direction, i.e., an item designating high self-esteem checked in the "Like Me" column or an item designating low self-esteem checked in the "Unlike Me" column was worth two points. The present paper reports only the total scores for the Self-Esteem Inventory.

The Inventory was administered by the experimenter, on a group basis, to two fifth- and sixth-grade classes. Preliminary instructions were designed to minimize anxiety, and indicated that the Inventory was not related to academic work, was not a test of achievement, and would remain confidential. The classroom teacher was not present during the administration. The instructions on the printed form were:

> Read each sentence carefully. If the sentence describes how you *usually* feel, put a check in the column "Like Me." If the sentence does not describe how you usually feel, put a check in the column "Unlike Me." There are no right or wrong answers.

Five weeks later the Inventory was readministered to one of the fifth-grade classes.

Scores on initial administration ranged from 40–100, with a mean of 82.3 and *SD* of 11.6. The mean score for the 44 boys was 81.3, *SD* of 12.2, and the mean score of the 43 girls was 83.3, *SD* of 16.7. The difference between the mean scores was not significant ($t = .80$; $p < .50$). The form of the distribution was skewed in the direction of high self-esteem. Test-retest reliability after a five-week interval with the sample of 30 fifth-grade children was .88.

*Self-Esteem Behavior Rating Form* (*BRF*). The teachers and the principal of the children involved in

[1] Copies of the Self-Esteem Inventory and Behavior Rating Form have been deposited with the American Documentation Institute. Order Document No. 5902 from ADI Auxiliary Publications Project, Photo Duplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

this study were asked to rate each child on a 14-item, five-point scale on behaviors presumed to be related to self-esteem. Items included in this rating schedule referred to such behaviors as the child's reactions to failure, self-confidence in a new situation, and the need for encouragement and reassurance. The behaviors to be rated were selected after a series of observations of behavior in and out of the classroom, repeated interviews with teachers, principals, and a clinical psychologist, and evaluations and discussions with a research committee.

The teacher and the principal rated the children independently and at no time collaborated in their ratings. The school and community situation was such that the principal had known most of the children for many years. The principal felt incapable of making precise ratings for 15 *S*s, and these were omitted from further consideration. Eight weeks after the first rating, one of the fifth-grade teachers was asked to again rate her class of 28 children.

The correlation between the evaluations of the teacher and the principal was .73 (Pearson product-moment). Since the correlation between the two ratings was high, further discussion of results is limited to the teacher ratings, which ranged from 23 to 100, with a mean of 68.4 and an *SD* of 15.4. The mean rating for the boys was 65.0, *SD*, 16.2; for the girls, 71.3, *SD*, 13.6. The difference between the means was significant ($t = 4.2$, $p < .001$). The test-retest reliability by one teacher of 28 children was .96.

### *Sample Selection*

Four groups of 12 children each were selected on the basis of the self-esteem inventory and the teacher and the principal ratings. Each group, representing various relationships between high-low self-evaluation and convergent-divergent behavioral evaluations, was composed of those 12 children who best met the criteria established. These criteria were as follows: (*a*) *High-High* (*H-H*)—scores on both the SEI and the Behavior Rating Form within the upper quartile of their class; (*b*) *Low-Low* (*L-L*)—scores on both the SEI and the BRF within the lowest quartile of their class; (*c*) *High-Low* (*H-L*)—SEI scores in the upper quartile and BRF scores in the lower quartile of the class; (*d*) *Low-High* (*L-H*)—SEI scores in the lower quartile and the BRF scores in the upper quartile of the class.

### *Academic, Social, and Personality Measures*

*Objective Achievement.* The Iowa Achievement Test score for each child was obtained from the school records. In the light of strong cultural pressures toward academic achievement, scholastic success and failure may be presumed to influence self-esteem.

*Sociogram.* The children in the four classes under study were asked to indicate which three children in their class they would most like to have as their friends. This information was compiled as the total number of times each child was chosen by his classmates. The measure was included on the assumption that the friendships and status of the child among his peers were related to his self-esteem.

*Childrens' Form of the Manifest Anxiety Scale.* Castaneda, McCandless, and Palermo (1956) have revised the Taylor Scale of Manifest Anxiety (1953) for use with fourth-, fifth-, and sixth-grade children. The scale consists of 42 items related to anxiety and 11 items included to provide an index of the *S*'s tendency to falsify his responses, and was included to ascertain whether differences in self-esteem would be related to differences in manifest anxiety.

*Achievement Motivation.* In an ego-oriented situation, the 48 *S*s selected according to the criteria previously noted were individually asked to complete four stories which were found by Winterbottom (1952)[3] to elicit achievement motivation in children. As the children told the stories they were written down by the *E*. The scoring of the stories followed the procedures and criteria presented by McClelland et al. (1953). A correlation of .85 (Pearson product-moment) was obtained between the scores of two judges on their scoring of the protocols.

*Ideal Self.* The 48 children selected for inclusion in the final sample were individually presented with the items in the Self-Esteem Inventory in card form and instructed to "Put these cards on either the 'Like Me' or the 'Unlike Me' pile according to the way you think the kind of person you would like to be would do it." The score for this Ideal Self was arrived at in the same manner as that for the Self-Esteem Inventory. This measure was included to ascertain any discrepancy between the *S*'s self evaluation and his ideal, which would presumably be an index of dissatisfaction with the present self picture.

## RESULTS

### *Success and Self-Esteem*

The scores on the achievement test and sociogram were used as indices of success experiences. The data were subjected to both correlational and chi square analysis, significant differences being obtained by both methods. Aspects of both analyses are presented here. In the chi square analysis, data were dichotomized at the class median. It appears likely (Hyman, 1942) that a student's evaluations of his performance are made within the smaller referent group of the class rather than the more generalized age status group. Table 1 shows the distribution of SEI scores in relation to academic and social success.

The chi square values of 5.1 for the achievement distribution and 8.3 for the sociometric choice distribution are both significant beyond the .05 level, and indicate a significant tendency for students above and below their class medians in success experi-

[3] I am indebted to Marian Winterbottom for her advice and cooperation with this phase of the study.

TABLE 1
ACADEMIC AND SOCIAL SUCCESS IN RELATION TO SELF-EVALUATION

| Self-Esteem | Iowa Achievement Test | | Sociogram | |
|---|---|---|---|---|
| | Above Class Median | Below Class Median | Above Class Median | Below Class Median |
| Above Class Median | 29 | 15 | 30 | 17 |
| Below Class Median | 18 | 25 | 13 | 27 |
| | $\chi^2 = 5.1$; $p < .05$ | | $\chi^2 = 8.3$; $p < .01$ | |

ences to be similarly above and below their class medians in self-esteem. It should be noted that although success is significantly related to self-esteem on both indices, there are many cases for whom this tendency does not hold true. These are scrutinized later in the paper.

Significant correlations were obtained between self-esteem and achievement ($r = .36$, $p < .01$), and between self-esteem and sociometric choice ($r = .37$, $p < .01$). Partial correlations between self-esteem and sociometric choice with achievement held constant ($r = .29$, $p < .01$), and between self-esteem and achievement with sociometric choice eliminated ($r = .30$, $p < .01$) indicated that self-esteem was independently related to these variables. The multiple correlation between sociometric choice and achievement combined ($r = .49$, $p < .001$) advances the prediction of self-esteem.

*Correspondence Between Evaluations*

Only in a minority of cases was there a marked discrepancy between the SEI and BRF scores. In 56 of the 87 cases, there was substantial agreement between the evaluations of the observer and the *S*. Substantial agreement was arbitrarily defined as a difference of less than 20 points, in either direction, between self-evaluation and the observer's rating of behavior. In only 8 of the 87 cases was there a difference of 40 points or greater. The distribution of differences between self- and observer ratings was positively skewed, with the 50th percentile falling at about 19, the 80th at about 31, and the 20th at about 3.8. The mean of the distribution was 18.7.

In order to ascertain whether the direction of the differences was consequential, the group showing the greatest amount of positive difference (High-Low) was compared with the group showing the greatest amount of negative difference (Low-High). Table 2 compares these groups on the available variables.

Inspection of Table 2 reveals that the two groups differ significantly on all but the Manifest Anxiety and Ideal Self indices. The differences in abilities, performances, and motivation clearly imply that positive (H-L) and negative (L-H) discrepancies do not represent similar personality types. Interpretations of the Ideal Self and the differences between Ideal and Self scores are tentative in the light of the possible ceiling effect. A further consideration is that the Self-Esteem, Ideal Self, and Manifest Anxiety scores are all based on self descriptions and are thus subject to a common response set. This set would result in high scores for the H-Ls and low scores for the L-Hs on all self-descriptions, a hypothesis that is not supported by the high Ideal Self score of the L-Hs and nonsignificant difference on the Manifest Anxiety Scale.

The high self-evaluation and high Ideal Self of the H-Ls in the face of poor achievement, low teacher rating, and low sociometric status would appear to indicate a defensive reaction. The alternative explanations of misperception or of generalization from a high degree of success in other areas of experience have not been examined in the present study.

The low self-evaluation of the L-Hs in the face of high achievement and sociometric status and of high teacher rating appears, in the light of the high Ideal Self and high Need Achievement scores, to represent a striving for high goals in the face of which the L-Hs present performance appears inadequate to them. The alternative explanations of misperception or of generalization from frequent failures in other areas of experience have again not been examined.

The cases of agreement between self- and observer evaluations remain to be examined.

TABLE 2
MEAN SCORES OF THE HIGH-LOW AND THE LOW-HIGH GROUPS

| Group | Self-Esteem Score | Teacher's Rating | No. of Times Chosen | Iowa Achievement | Manifest Anxiety/Lie | Ideal Self | Ideal Self Difference | Need Achievement |
|---|---|---|---|---|---|---|---|---|
| High-Low | 90.0** | 48.9** | 2.4** | 5.8** | 9.5/4.9 | 91.3 | 1.3** | 6.3* |
| Low-High | 73.1 | 97.1 | 4.6 | 8.5 | 15.1/3.2 | 92.2 | 19.1 | 8.8 |

* Difference between scores significant at .05 level.
** Difference between scores significant at .01 level.

TABLE 3
MEAN SCORES OF THE HIGH-HIGH AND THE LOW-LOW GROUPS

| Group | Self-Esteem Score | Teacher's Rating | No. of Times Chosen | Iowa Achievement | Manifest Anxiety/Lie | Ideal Self | Ideal Self Difference | Need Achievement |
|---|---|---|---|---|---|---|---|---|
| High-High | 93.1** | 96.0** | 4.1** | 7.3** | 4.5*/4.5 | 94.4* | 1.3* | 4.2 |
| Low-Low | 71.6 | 30.8 | 1.8 | 5.4 | 12.0/4.7 | 81.9 | 10.3 | 4.1 |

* Difference between scores significant at .05 level.
** Difference between scores significant at .01 level.

Table 3 compares the scores of the High-High group with the Low-Low group on the variables for which information was collected. These two groups, comprised of individuals who best met the criteria established, are clearly not a random sample of all cases of agreement between self- and observer ratings, since there were cases of agreement along the entire distribution. They do, however, indicate that agreement of self- and observer evaluation is not in itself an indication of similarity on these variables.

Table 3 reveals that the High-Highs and the Low-Lows differ significantly on all variables but Need Achievement. The implication is that agreement of subjective evaluation and behavioral expression does not in itself indicate similarity of personality types. The possibilities of a ceiling effect and response set are again present and as was true for the L-H and H-L groups, the teacher's ratings and students' sociometric choices are in agreement, the H-Hs being rated favorably by both groups and the L-Ls unfavorably.

*Types of Self-Esteem*

The results noted in Tables 2 and 3 imply that profiles of the characteristics of the children in the different groups differ markedly. Such profiles consolidate the results of the various measures to bring out the psychological significance of each variable in context.

*Low-High*—Children in this group are notable for their extremely low self-evaluation in the face of high teacher evaluation, high sociometric status, and extremely high Iowa Achievement Test scores. In the face of such overwhelming evidence of superiority, why the low self-evaluation? To follow Winterbottom's (1952) suggestions, it would appear that L-Hs are competing against absolute standards of excellence rather than against the performance of their peers. The alternative possibility that L-Hs do not perceive the high esteem in which they are held must remain secondary when one considers the continual reinforcement in achievement, peer, and teacher recognition that they receive. That these children have high goals against which current accomplishments seem unimportant is borne out by a high Need Achievement score, a high Manifest Anxiety score, and a high D (Ideal Self-Self) score. Performance is relative to the goal, and the goals of L-Hs are apparently set so high that despite their high level of performance and acceptance by others they see themselves as performing poorly and, as a result, are anxious and hard-driving. The foregoing profile appears to characterize the *S*s above the median in success and below the median in self-esteem in Table 1.

*High-Low*—Children in this group showed a high self-evaluation although they were

rated low by the teacher and were seldom selected as friends by their classmates. Their high self-evaluation in the face of the low ratings by others and their low academic achievement appears to represent a defensive reaction. Seen in this perspective, the relatively low Manifest Anxiety score of the H-Ls, with an L score higher than that of any other group, suggests a need to defend against conscious awareness of their low status. Although the scores on the Manifest Anxiety and L scales were in the expected direction for all groups, the differences are not significant. The high Ideal Self and Need Achievement scores corroborate the view that the H-Ls are unwilling to accept their low status and are striving for higher goals. The H-L's responses indicative of high self-esteem are based neither on performance nor on positive opinions toward him of those in the classroom, but are maintained and defended in the absence of status and accomplishment. This profile fits the *Ss* below the median in success and above the mean in self-esteem.

*Low-Low*—Children in this group had the lowest mean self-evaluation of any group. Their behavior was evaluated as lowest by the teacher, they were the least chosen as friends by their peers, and they had the lowest scores on the Iowa Achievement test. The assumption that they recognize their low status is borne out by the relatively high Manifest Anxiety score (despite a high L score) and the low Ideal Self they profess. The latter is taken as an indication that the L-Ls recognize their low status but do not set high goals, since past attempts to improve their position have proved painful and useless. Such an interpretation is consistent with their low achievement motivation score. The relatively large D (Ideal Self-Self) score, evaluated in the context of self and Ideal Self scores, may indicate either a nonacceptance of their status, or more likely, a high verbalized ideal but no behaviors that indicate active seeking for this ideal. L-L children presumably have learned that they can function best by accepting a low status.

*High-High*—The children in this group had the highest mean self-evaluation of any group. They were rated highly by themselves, the teacher, and were frequently chosen as friends. The small D (Ideal Self-Self) score suggests they are content with themselves, a finding supported by their low Need Achievement and low Manifest Anxiety scores. Achievement test scores indicate they are only slightly above average for their classes. H-Hs, then, have a high opinion of themselves, are highly thought of by others, and express their opinions in their behaviors. Though not the highest achievers in the class, the fact does not appear to disturb them, and they seem content with the status quo.

## DISCUSSION

A person's self-evaluation thus generally agrees with those of his behaviors which, on either theoretical or empirical grounds, presumably reflect that self-evaluation. Where there is agreement there appears to be a close congruence between ideational and behavioral expression so that changes in the one are reflected in the other. For those cases where there is not agreement, the findings suggest the operation of two mechanisms. In the first instance, high manifest self-evaluation apparently is a reaction against a low underlying self-evaluation as a result of poor performance and low status. The second is marked by low self-evaluation despite excellent performance and high status, and appears due to a striving for extremely high goals. Although differing in many respects, both types of disagreement appear to be similar in their refusal to accept the status quo and are marked by high achievement motivation, a high level of anxiety (manifest or latent), and an ideal beyond immediate reach.

The study raises questions concerning the definition of adjustment and self-esteem, and the methods by which they are determined. Thus, by a generally accepted definition of adjustment (Barron, 1953), "getting along in the world as it is, adequate degree of social conformity, capacity to adapt to a wide range of conditions, ability to fit in" (p. 166), it is the H-Ls and L-Ls who are maladjusted. According to definitions of adjustment that emphasize accuracy of perception, however, it is the H-Ls and L-Hs who are maladjusted. Furthermore, the H-Hs, who would appear to be the ideal of all adjustment criteria, do not

appear to be as forceful, constructive, complex, or as potentially creative as do the L-Hs. The H-Ls would appear to be the group to which the criteria of maladjustment would most generally be applied. Yet even in these cases there is reason to question the appropriateness of its application, unless adjustment is defined solely in terms of such concepts as passivity, acceptance, and maintenance of the status quo.

Two major impressions regarding the definition of self-esteem emerge from the present study. The first is that a clarification and definition of what is meant by self-esteem is essential, and the second, intimately related to the first, is that self-esteem is an ephemeral subject difficult to deal with empirically. Thus when self-esteem is defined solely by subjective evaluation, defensive processes render measures so based ambiguous, whereas a definition solely from the perspective of the observer overlooks a vital component. When both self- and observer evaluations are used, the result is a series of types based upon level of evaluation and extent of agreement between evaluations. Which of all the foregoing is self-esteem? In the final analysis it must be admitted that they all are. There is thus the self-esteem the individual purports to have, the self-esteem which he subjectively holds, the self-esteem he displays (or attempts to display), and the self-esteem behavior that is observed and reported by others. The self-esteem behavior that the individual displays is presumably based, to a great extent, upon his prior positive and negative experiences. His "real" self-esteem can only be inferred from phenomenal reports, evaluation of the unconscious, and behavioral evaluations, and may be expected to vary with the structure and demands of a given situation. Similarly, unless so demonstrated, it cannot be assumed to remain identical for different areas of experience, or even for the same area of experience over a period of time. The choice of any particular self-esteem as "real" would appear to be valid to the extent that it increases understanding and prediction of the specific behavior in question. It would thus seem, at the present stage of psychological knowledge, that a combination of subjective and observer evaluations, taken in conjunction with a constellation of variables relevant to the major variable in question, would appear to offer a promising method for avoiding the pitfalls attendant to the use of any one of these methods separately. Greater predictability may be expected, since the experimenter can then determine which combination of self-esteem variables is most pertinent to a given situation.

The present instruments for self-evaluation and behavior rating clearly involve selection from a potentially wide range. The high reliability of both the self- and the behavioral evaluations nevertheless indicate that similar and consistent interpretations of both measures were made by a large number of individuals. Subsequent reports on this research will indicate the extent to which differences in self-esteem determine behavior in a variety of experimental situations.

## SUMMARY

The present study describes a method of determining self-esteem. This method is based upon a combination of (*a*) the subject's response to the Self-Esteem Inventory; (*b*) a rating of certain of his behaviors which are presumably related to self-esteem; and (*c*) information on a constellation of experiential and motivational variables.

For a group of 10–12-year-old children, there was substantial agreement between self-evaluation and behavioral expression in a majority of cases. This agreement was expected on the grounds that reality experiences intrude upon the individual to such an extent that they generally cannot be avoided. It was in fact found that persons who had more success experiences were significantly higher in their self-evaluation than individuals with fewer such experiences.

High test-retest reliability was found for the Self-Esteem Inventory and the behavioral rating scale. Four groups of 12 subjects each were selected to represent the four possible patterns of extreme standing on the two variables taken jointly. The four groups were found to differ significantly in achievement, sociometric status, ideal self, and achievement motivation and, apparently, represent distinct types of self-esteem. The groups in

which self-evaluation disagreed with behavioral evaluation were significantly higher in achievement motivation than those which showed agreement.

It is suggested that attempts to determine self-esteem be broadened to include behavioral, phenomenal, and experiential data.

## REFERENCES


ALPER, THELMA G. Predicting the direction of selective recall: Its relation to ego strength and *n* Achievement. *J. abnorm. soc. Psychol.*, 1957, **55**, 149–165.

ASCH, S. The doctrine of suggestion, prestige, and imitation in social psychology. *Psychol. Rev.*, 1948, **55**, 250–85.

ATKINSON, J. W. Achievement motivation and the recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

BALDWIN, A. L., & LEVIN, H. Effects of public or private success or failure on children's repetitive motor behavior. Unpublished manuscript, Cornell Univer., 1957.

BARRON, F. Complexity-simplicity as a personality dimension. *J. abnorm. soc. Psychol.*, 1953, **48**, 163–173.

CASTENADA, A., MCCANDLESS, B., & PALERMO, D. The children's form of the Manifest Anxiety Scale. *Child Developm.*, 1956, **27**, 327–333.

ERIKSON, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

FREUD, ANNA. *The ego and the mechanisms of defense.* London: Hogarth, 1937.

FREUD, S. *The ego and the id.* London: Hogarth, 1927.

GOUGH, H. G. A preliminary guide for the use and interpretation of the California Psychological Inventory. *Res. Bull.*, Inst. of Personality Assessment and Research, Berkeley, 1954.

HORNEY, KAREN., *The neurotic personality of our time.* New York: Horton, 1937.

HYMAN, H. H., The psychology of status. *Arch. Psychol., N. Y.*, 1942, No. 269.

JOURARD, S. M. Identification, parent cathexis, and self-esteem. *J. consult. Psychol.*, 1957, **21**, 375–380.

LAZARUS, R. S., & LONGO, H. The consistency of psychological defenses against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495–499.

MASLOW, A. H. Dominance, personality and social behavior in women. *J. soc. Psychol.*, 1939, **10**, 3–39.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A. & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

ROGERS, C. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.

ROGERS, C., & DYMOND, R. *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

ROSENZWEIG, S. The experimental study of repression. In H. A. Murray (Ed.), *Explorations in personality.* New York: Oxford Univer. Press, 1938.

SEARS, PAULINE S. Levels of aspiration in academically successful and unsuccessful children. *J. abnorm. soc. Psychol.*, 1940, **35**, 498–536.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–289.

WINTERBOTTOM, MARIAN. Relation of childhood training in independence to achievement motivation. Unpublished doctoral dissertation, Univer. of Mich., 1952.


Received May 1, 1958.

# CONFLICT AND REPRESSION DURING SLEEP DEPRIVATION

EDWARD J. MURRAY[1]

*Walter Reed Army Institute of Research*

THE purpose of this study was to examine adjustment to a conflict generated by prolonged sleep deprivation with the aid of a modified Thematic Apperception Test. The conflict appeared to be between the physiological need for sleep and a number of social pressures to stay awake. Previously reported research (Murray, Schein, Erikson, Hill, & Cohen, in press; Murray, Williams, & Lubin, 1958) suggested that repression, or a repression-like process, was a primary mode of adjustment. Therefore, this study was designed to measure a repressive process.

The concept of repression was used by Freud (1936) to account for difficulties in recalling anxiety-producing material of a sexual or aggressive nature. However, Dollard and Miller (1950), in their behavioral analysis of repression, have extended the concept to additional classes of drives and responses. They distinguish three types of repression of a drive: the inhibition of responses labeling the drive, the inhibition of responses producing the drive, and the inhibition of responses mediating the drive. In the case of the sleep drive, an example of the inhibition of responses labeling the drive might be the finding of Murray, Williams, and Lubin (1958) that sleep-deprived subjects tended to deny feelings of sleepiness. During 98 hours of sleep loss these subjects rated themselves significantly lower on a sleepiness scale than did staff observers. The discrepancy increased significantly from day to day up until the last few hours of the experiment when the subjects' ratings suddenly increased dramatically. Self-ratings of fatigue, which appeared to be less ego-involving, were significantly higher than the self-ratings of sleepiness. An example of the inhibition of responses producing the drive might be the finding of Murray, Schein, Erikson, Hill, and Cohen (in press) that subjects deprived of sleep for 98 hours tended to change restlessly from one form of social and recreational behavior to another. This was interpreted as an avoidance of the overpowering drowsiness that would occur when the subjects persisted in any one activity for any length of time. The Dollard and Miller theory of repression would also suggest that responses mediating the sleep drive would also be inhibited. Such mediating responses might include the overt and covert verbalizations that are involved in TAT fantasy productions. Thinking about sleep might increase the tendency to go to sleep, while avoiding sleep thoughts would help maintain wakefulness. Therefore, according to this theory of repression, TAT sleep themes should decrease during sleep deprivation.

During the past several years there have been a large number of studies concerned with the effects of motivational variables on perceptual, associative, and imaginal processes (e.g., Granger, 1953; McClelland, Atkinson, Clark, & Lowell, 1953; Spence, 1957). Although the evidence is fragmentary and not fully consistent, there is some indication that drives such as hunger produce a perceptual, associative, and imaginal sensitization at low and moderate levels of drive but activate a defensive process at higher levels (Lazarus, Yousem, & Arenberg, 1953; Wispe, 1954). Eriksen and Browne (1956) have made a behavioral analysis of this kind of defense in terms of avoidance of anxiety-producing verbal responses. On the basis of the evidence from this area of research, one might also expect a decrease in TAT sleep themes after a prolonged period of sleep deprivation.

## METHOD

### *Subjects and Setting*

The subjects (*Ss*) were 48 soldiers, including 24 experimentals and 24 controls, who had volunteered for a larger study of the effect of sleep deprivation upon a number of physiological and psychological variables described more fully elsewhere (Murray et al., in press). The *Ss* reported in groups of 10 for a 12-day period.

[1] The author wishes to thank Melvin Cohen for helping with some of the statistical computations and the people who helped with the ratings. The author also wishes to thank David C. McClelland and Ardie Lubin for critically reading the manuscript and making many valuable suggestions. The author is now at Syracuse University.

Each group was divided into five experimental and five control *Ss* matched for intelligence. After four preliminary days, the experimentals were kept awake for 98 hours and then observed for four more days. The experimentals were housed on a hospital ward and kept awake by gentle means. The controls slept on a nearby ward but were with the experimentals during the day. *Ss* were naive with respect to the TAT and the repression hypothesis. All *Ss* successfully completed the experiment.

### *Sleep and Fatigue Themes*

In studying TAT sleep themes it was decided to vary the ambiguity of the TAT stimulus cards with respect to degree of sleep suggestion, since previous research and comment had indicated the importance of this variable (Epstein & Smith, 1956; Lazarus, 1953). Three TAT cards representing a continuum of sleep suggestion were selected in the following way: six TAT cards from the standard set (Murray, 1938) were selected by the author as covering the spectrum of sleep suggestion. These six cards (1, 3, 9 GF, 9 BM, 14, and 17 BM) were then ranked by 10 doctoral level psychologists according to degree of sleep suggestion. The three cards with the average rank of 1.0, 3.0, and 5.6 were selected to represent high, medium, and low sleep suggestion. These three cards were then given to 10 secretaries, assistants, and soldiers to rank according to degree of sleep suggestion. There was 100% agreement on this ranking. The final sleep series included: a picture of several hobos lying around (9 BM, high sleep suggestion); a picture of a boy sitting and looking at a violin (1, medium sleep suggestion); and a man climbing a rope (17 BM, low sleep suggestion).

The stories to these TAT cards were analyzed for sleep themes according to a simple scoring system based on behavioral concepts (Dollard & Miller, 1950) and modelled after the McClelland systems for hunger, achievement, power, sex, affiliation, and other needs (Atkinson & McClelland, 1948; Atkinson, Heyns, & Veroff, 1954; Clark, 1952; McClelland et al., 1953; Veroff, 1957). First, the presence or absence of a sleep theme in general was noted. Then the presence or absence of the following subthemes was determined: (*a*) conditions producing sleep drive, including going without sleep, eating a full meal, sitting in a stuffy room; (*b*) subjective feelings of sleep drive, including descriptions of sleepiness, drowsiness, or grogginess; (*c*) instrumental responses for reducing sleep drive, including looking for an opportunity to sleep, making attempts to get to sleep, or lying down to sleep; and (*d*) goal responses reducing sleep drive, including actual sleeping behavior.

A similar scoring system was devised for fatigue themes. The drive in general and the following drive-components were rated: (*a*) conditions producing fatigue drive; (*b*) subjective feelings of fatigue drive; (*c*) instrumental techniques for reducing fatigue drive; and (*d*) goal behavior reducing fatigue drive.

The stories were coded and scored in a random order by an experienced clinical psychologist not aware of the purpose of the study. A second clinical psychologist was used for a reliability check, which showed from 92 to 99% agreement on each of the 10 categories. These percentages were all significant at the .01 level except for feelings of fatigue (.07) and conditions producing sleep (few instances, not significant).

### *Hostility Ratings*

Three TAT cards representing a continuum of overt hostility suggestion were selected in a procedure similar to that used with the sleep series. Five regular TAT cards (2, 4, 7 GF, 10, and 18 GF) and one especially designed card (a western scene showing two men fighting) were ranked by 10 doctoral psychologists according to degree of overt hostility suggestion. Three cards (special, 4, and 10) with average ranks of 1.1, 2.8, and 5.9 were then ranked by 10 secretaries, assistants, and soldiers with 100% agreement.

Global ratings of presence or absence of overt hostility were made on each of the stories by a different clinical psychologist. Still another clinical psychologist was used for a reliability check, which showed a 78% agreement ($p > .001$.)

### *Tension Ratings*

Each of the stories to the three sleep cards and the three hostility cards was rated globally for tension by another clinical psychologist. Tension was defined in the broadest sense to include anxiety, frustration, high drive, etc. A psychological assistant was used for a reliability check, which showed 78% agreement, significant at the .01 level.

### *Word Counts*

The number of words in each story given in response to the three sleep cards and the three hostility cards was counted by a dependable assistant.

### *Eron's Categories*

In a normative study of the TAT Eron (1950) defines a number of empirically derived content categories including: ***retirement***, which appears to be similar to a sleep theme; and ***exhaustion***, which appears to be similar to a fatigue theme. In order to make a comparison with a group of *Ss* not involved in the sleep deprivation study at all, the stories of control and experimental *Ss* in response to Card 9 BM (high sleep suggestion) were scored in terms of 13 categories found useful in analyzing the content of stories in response to this card by a group of 75 nonhospitalized *Ss* (college students and psychoneurotics) in Eron's study. These categories are listed in Table 4 below. A story could have more than one content theme. A clinical psychologist rated all the stories in response to Card 9 BM by control and experimental *Ss*, and another psychologist was used for a reliability check. Agreement on the presence or absence of the theme ranged from 79 to 100%. These percentages were significant at the .01 level except for economic pressure, general and legal restriction, and the infrequently occurring category of aggression from environment. The key category, retirement, showed 92% agreement, significant at the .01 level.

### *Administration Procedure*

The *Ss* were awake from 6:00 on a Friday morning until 8:00 on a Tuesday morning. The TAT was administered between about 8:00 and 12:00 on the Monday night or at from 86 to 90 hours of sleep deprivation. During each cycle of the experiment, five experimental and five controls were given the six TAT cards described above. The experimental and control *Ss* were alternated during the evening. The *Ss* were asked to dictate a story in the usual way (Murray et al., 1938), but probing was kept to a bare minimum, a procedure that resulted in somewhat short stories. Two men were omitted from the last of the five cycles, leaving a total of 24 experimental and 24 control *Ss*. Half of each group had the sleep and hostility series presented in an ascending order (e.g., low, medium, and high sleep suggestion) and half in a descending order (e.g., high, medium, and low sleep suggestion). In turn, half of each of these groups had the sleep series first, and half had the hostility series first. During the analysis of the ratings it was found that neither the order of presentation, the order of the series, nor the number of the cycle had any demonstrable effect on the sleep, fatigue, or other ratings.

## RESULTS

The results as a whole indicate that sleep deprivation tends to decrease the thematic expression of sleep needs. This effect is shown in Fig. 1 where it can be seen that a larger percentage of control *Ss* than experimental *Ss* wrote TAT stories containing sleep themes. This difference is significant ($p < .01$, by Fisher's exact test) for the high sleep suggestion card. The difference on the medium sleep suggestion card is in the predicted direction,

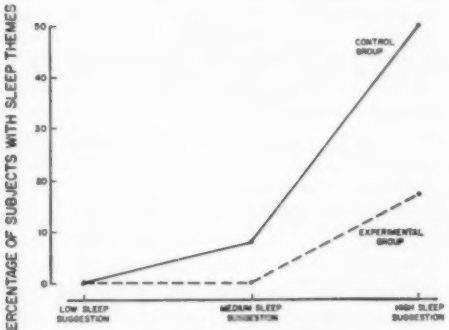


FIG. 1. PERCENTAGE OF *Ss* IN THE SLEEP-DEPRIVED AND CONTROL GROUPS WITH SLEEP THEMES PRESENT IN STORIES WRITTEN IN RESPONSE TO TAT CARDS REPRESENTING A CONTINUUM OF SLEEP SUGGESTION

TABLE 1
SLEEP SUBTHEMES ON HIGH SLEEP CARD

| Sleep Subtheme | Percentage of Control Group | Percentage of Experimental Group |
|---|---|---|
| Conditions producing sleepiness | 4 | 0 |
| Feelings of sleepiness | 4 | 8 |
| Instrumental responses | 29 | 8 |
| Goal responses | 42 | 4 |

although not significant, while there is no difference on the low sleep suggestion card.

Sleep deprivation decreases not only the occurrence of over-all sleep themes but subthemes as well. Table 1 shows the sleep subthemes in response to the high sleep suggestion card, which was the only card with enough sleep themes in the experimental group to justify detailed analysis. A smaller percentage of the experimental *Ss* than control *Ss* wrote stories in which the subtheme of goal responses appeared ($p < .02$, Fisher's exact test). The instrumental response subtheme was less prevalent in the experimental than the control group but at a lower level of significance ($p < .06$, Fisher's exact test). Conditions producing sleepiness and feelings of sleepiness subthemes did not occur with much frequency in either group and showed no significant differences between groups. It would appear that sleep deprivation decreases all discernible sleep subthemes.

To what extent do other over-all themes decrease with sleep deprivation? Table 2 compares the occurrence of over-all sleep themes with fatigue themes, tension ratings, and average number of words in control and experimental groups in response to the three sleep suggestion cards. The percentage of fatigue themes is related to the degree of sleep suggestion in the stimulus cards but shows no significant differences between experimental and control groups. Tension ratings show no differences between experimental and control groups but do show a slight decrease on the high sleep suggestion card. The average number of words per story shows no systematic differences between cards but tends to be slightly higher for experimental *Ss*. Sleep is the only theme in response to the sleep cards

TABLE 2
THEMES IN RESPONSE TO SLEEP CARDS

| Theme or Measure | Control Group | | | Experimental Group | | |
|---|---|---|---|---|---|---|
| | Degree of Suggestion | | | Degree of Suggestion | | |
| | Low | Medium | High | Low | Medium | High |
| Percentage of *Ss* with sleep theme | 0 | 8 | 50 | 0 | 0 | 17 |
| Percentage of *Ss* with fatigue theme | 8 | 0 | 50 | 8 | 4 | 54 |
| Percentage of *Ss* with high tension rating | 62 | 62 | 29 | 67 | 58 | 37 |
| Average number of words per story | 38 | 40 | 38 | 58 | 49 | 49 |

TABLE 3
THEMES IN RESPONSE TO HOSTILITY CARDS

| Theme or Measure | Control Group | | | Experimental Group | | |
|---|---|---|---|---|---|---|
| | Degree of Suggestion | | | Degree of Suggestion | | |
| | Low | Medium | High | Low | Medium | High |
| Percentage of *Ss* with hostility theme | 8 | 51 | 88 | 8 | 54 | 79 |
| Percentage of *Ss* with high tension ratings | 46 | 87 | 50 | 25 | 83 | 46 |
| Average number of words per story | 37 | 40 | 35 | 65 | 70 | 50 |

that shows a significant difference between experimental and control groups.

Hostility is another theme that might be presumed to show a difference, one way or the other, between the control and sleep-deprived groups. Table 3 shows hostility themes along with tension ratings and average number of words per story for both experimental and control *Ss*' stories in response to the three hostility suggestion cards. The percentage of hostility themes is related to the degree of hostility suggestion of the cards but shows no significant differences between experimental and control groups. Tension ratings show no differences between groups, although they tend to be higher for the medium hostility suggestion card for some unclear reason. The average number of words per story seems about the same for all cards but tends to be slightly higher for the experimental group. This slight difference does not appear to be systematically related to any other variable. In general, the experimental and control groups do not differ with respect to hostility themes.

Although the control *Ss* slept normally throughout the sleep deprivation period they were intimately involved in the project as a whole and may have given somewhat atypical stories. One possible means of controlling for this is to compare the stories of the control and experimental *Ss* with normative data. Table 4 shows the percentage of a group in which a particular theme occurred in the study in response to the high sleep suggestion card (9 BM). The table compares Eron's normative group with the present control and experimental groups on Eron's 13 empirically derived categories. The over-all pattern of themes is similar, but only the first three themes listed will be considered because the others either do not occur very frequently or are not reliable. The retirement theme (sleep) occurs significantly more often in the control group than the experimental and in the normative group more often than either the experimental or control ($p < .05$ or better, by a one-tailed chi-square test). The exhaustion (fatigue) is significantly higher in both the control and experimental groups than in the normative group ($p < .01$). The exhaustion theme occurs more frequently in the sleep-deprived than the control group, but this difference is not significant. The only other significant difference is between the normative and control group on vacillation ($p < .01$).

TABLE 4
PERCENTAGES OF NORMATIVE, CONTROL, AND SLEEP-DEPRIVED GROUPS WITH EMPIRICAL THEMES TO HIGH SLEEP CARD

| Empirical Theme | Normative Group | Control Group | Experimental Group |
|---|---|---|---|
| Retirement (i.e., sleep) | 68 | 42 | 12 |
| Exhaustion (i.e. fatigue) | 15 | 42 | 62 |
| Vacillation | 13 | 50 | 25 |
| Tranquility | 16 | 0 | 4 |
| Economic Pressure | 9 | 0 | 4 |
| War | 9 | 4 | 12 |
| Congeniality-Peers | 8 | 4 | 0 |
| General Restriction | 7 | 17 | 17 |
| Succorance-Peers | 4 | 4 | 0 |
| Reminiscence-Happy | 4 | 0 | 0 |
| Legal Restriction | 3 | 4 | 0 |
| Aggression-From Environment | 3 | 12 | 25 |
| Drunkenness | 1 | 8 | 8 |

These results are quite comparable to those already presented and indicate that the difference in sleep themes between the experimental and control groups is due to a decrease in the experimental and not an increase in the control group.

## DISCUSSION

The results show that sleep deprivation tends to decrease the frequency of occurrence of sleep themes in stories in response to TAT stimulus cards. This decrease is found with both instrumental and goal response subthemes and probably extends to other subthemes of sleep. The decrease may be viewed as an increase in the threshold for sleep themes in response to TAT stimuli since, as the sleep suggestion of the cards increases, some sleep themes can be evoked although not as many as with the control group. The variation in frequency of sleep themes with the sleep suggestion of the cards in both the experimental and control groups supports the validity of the ranking method used to establish the stimulus series. The decrease in sleep themes does not appear to be an artifact of a general inhibitory effect of sleep deprivation, since fatigue themes, word counts, hostility ratings, and tension ratings showed no decreases with sleep loss. On the other hand, the fatigue themes and hostility ratings showed meaningful relationships with aspects of the stimulus cards indicating some validity for these measures. The reliability of all of the measures appears satisfactory. The comparison of the present results with normative data shows that the lower frequency of sleep themes in the sleep-deprived group is not due to an unusual increase in the control group. There is some indication that the control group is in between the sleep-deprived and normative groups in frequency of sleep themes. This relationship may possibly be due to an identification with the experimental *S*s by the control *S*s. The comparison with the normative group also raises the possibility that both the experimental and control groups, but particularly the experimental group, may be expressing sleep needs in the displaced form of fatigue needs (Miller, 1948; Murray & Berkun, 1955). However, in view of the lack of corroborating evidence this possibility must remain conjectural.

The results fit in with previous research in a number of ways. High levels of the sleep drive lead to a defensive process similar to that shown in response to high levels of hunger (Lazarus et al., 1953; Wispe, 1954). However, reports of a differential effect of high drive on instrumental and goal response categories are not supported by these data. But it is unlikely that these studies involving no more than 24 hours of food deprivation were dealing with drive levels comparable in strength to 86 hours of sleep deprivation. In the one study involving hunger comparable to or greater than the sleep drive used in this study, it was also found that most projective indices of hunger showed no increase and that subjective ratings of hunger were probably underestimates (Brozek, Guetzkow, & Baldwin, 1951). The present results fit in with earlier research indicating an avoidance of self-descriptions or situations which would increase sleep needs (Murray et al., in press; Murray, 1958). The fact that fatigue themes do not decrease with sleep deprivation is comparable with the earlier finding that self-ratings of fatigue were higher than those of sleep. The present results support the earlier work, and comment suggesting that the ambiguity of the stimulus card is important (Epstein & Smith, 1956; Lazarus, 1953). But the results do not support the finding that displacement may be operating to increase the need-related themes to the cards least suggesting the need while at the same time decreasing need-related themes to the cards most directly suggesting the need (Epstein & Smith, 1956). As mentioned earlier, there may be some displacement from sleep to fatigue needs in this study. In evaluating this discrepancy two factors have to be considered. First, the drive level employed in this study may have been so high as to generalize the inhibition of the sleep need to all TAT-like stimuli (Dollard & Miller, 1950). Second, the TAT stimuli in the two studies may have represented quite different places in the continua involved; this possibility would require an extensive psychophysical study to evaluate. Again, the discrepancy is most likely due to the fact that the drive level used in the present study is considerably higher than that employed in most previous studies on the relationship be-

tween motivation and perceptual, associative or imaginal processes.

The nearly complete absence of hostility in the projective material or in the behavioral observations seems to contradict earlier studies in which it was found that sleep deprivation acts as a frustration and produces aggression (Sears, Hovland, & Miller, 1940) and psychotic-like delusions and hallucinations (Tyler, 1947). However, it must be remembered that in the present study every effort was made to be friendly and courteous to the *S*s. They were awakened by gentle means if they fell asleep, although occasionally even this procedure produced some mild irritation if they were just on the verge of falling asleep. On the other hand, these earlier studies involved a number of other frustrations including broken promises, not permitting smoking, etc. in the one study (Sears et al., 1940) and severe military discipline, periodic marches, etc. in the other (Tyler, 1947). Therefore, these hostile and psychotic reactions probably resulted from a complex, frustrating social situation and not necessarily from sleep loss per se.

The results of the present study along with the previous ones in this series (Murray et al., in press; Murray, 1958) may be interpreted in terms of the Dollard and Miller (1950) learning analysis of suppression and repression. Sleep may be construed as a physiological drive with a number of learned components. These learned components might include: verbal responses labeling the sleep drive, corresponding to the self-ratings of sleepiness; responses producing an increase in the sleep drive, corresponding to the responses involved in persisting in activities like hobbies; and responses mediating the sleep drive, corresponding to the complex responses involved in imagining and telling a TAT story. All of these responses tend to increase the sleep drive by adding to the physiological processes stimulating the person to go to sleep. During the sleep deprivation period the response of going to sleep was in strong conflict with a number of social pressures. The exact nature of these social pressures is not known, but they would appear on the basis of qualitative observation to include the *S*'s needs to appear strong, masculine, and grown-up in comparison with his fellows and in the eyes of the staff, as well as fears of failure, shame, and humiliation. (An example of this was the soldier who at the outset seemed likely to "goof off" at the earliest possible moment but who voluntarily endured a painfully abscessed tooth in order to complete the experiment successfully.) In any case, the *S*s were in conflict, and there is every reason to believe that the conflicting tendencies were closely balanced. The *S*s adjusted to this situation by avoiding those responses which would add to the already strong physiological stimulus for going to sleep: the sleep-labeling, sleep-evoking, and sleep-mediating responses involved in self-rating, social behavior, and thematic apperception. It is difficult to say to what extent this was unconscious repression or conscious suppression. The qualitative evidence would suggest that the *S*s were not aware of avoiding sleep themes on the TAT. Similarly, the self-ratings were done with a somber seriousness. On the other hand, at least a few of the *S*s were aware of avoiding feelings of drowsiness in their restless changing of social activities. In general, the behavior appeared to involve a mild repression or what in psychoanalytic parlance might be termed "preconscious phenomena." Since repression usually refers to some sort of verbal behavior, it might be more accurate to refer to the activity change as gross behavioral avoidance. But since it functions in a manner so comparable to the verbal responses it is clear that the two forms of behavior are very similar. Therefore, the results can be viewed as aspects of a mild repression of sleep needs as a means of maintaining wakefulness during prolonged sleep deprivation.

## SUMMARY

1. A series of three TAT cards representing a continuum of low, medium, and high sleep suggestion was selected on the basis of rankings by 10 psychologists and 10 lay people. These three cards were administered in the usual TAT way to 24 experimental *S*s after about 86 hours of sleep deprivation, as well as to 24 control subjects. Another series of three TAT cards representing low, medium, and high hostility suggestion devised in a similar way was administered at the same time. Half of each group was given the two series in an ascending order and half in a descending

order, while half of each of these groups was given the sleep series first.

2. The main result was that the sleep-deprived group gave fewer sleep themes than did the control group. Neither group had sleep themes on the low sleep suggestion card, the control group had a few on the medium card, and the control group had significantly more on the high sleep suggestion card. This decrease with sleep deprivation was true not only for the over-all sleep theme but also for subthemes such as instrumental response and goal response. Scoring for fatigue themes or tension failed to discriminate between the groups. There was no difference in hostility themes on the hostility series. Order of presentation was not significant. Scoring was reliable. A comparison with Eron's normative data for themes on the high sleep suggestion card suggested that the control group was in between the normative group and the experimental group on several relevant variables.

3. The results of this study were interpreted as indicating that the subjects were avoiding thoughts or verbal responses which might arouse overpowering sleepiness. These results were compared with previous studies in this series which suggested that the *Ss* denied sleepiness in self-ratings and avoided sustained activities which would produce sleepiness. An analysis of these phenomena in terms of the Dollard-Miller theory of repression was made. It was concluded that the results represent the use of a mild repression as a way of coping with a conflict between the sleep drive and various social pressures to stay awake.

## REFERENCES


ATKINSON, J. W., & MCCLELLAND, D. C. The projective expression of needs: II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.

ATKINSON, J. W., HEYNS, R. W., & VERHOFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

BROZEK, J., GUETZKOW, H., & BALDWIN, MARCELLA. A quantitative study of perception and association in experimental semi-starvation. *J. Pers.*, 1951, **19**, 245–264.

CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

EPSTEIN, S., & SMITH, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.*, 1956, **20**, 372–384.

ERIKSEN, C. W., & BROWNE, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224–230.

ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).

FREUD, S. *The problem of anxiety.* New York: Norton, 1936.

GRANGER, G. W. Personality and visual perception: A review. *J. ment. Sci.*, 1953, **99**, 8–43.

LAZARUS, R. S. Ambiguity and nonambiguity in projective testing. *J. abnorm. soc. Psychol.*, 1953, **48**, 443–445.

LAZARUS, R. S., YOUSEM, H., & ARENBERG, A. Hunger and perception. *J. Pers.*, 1953, **21**, 312–328.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MURRAY, E. J., & BERKUN, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, **51**, 47–56.

MURRAY, E. J., SCHEIN, E. H., ERIKSON, K. T., HILL, W. F., & COHEN, M. The effects of sleep deprivation on social behavior. *J. soc. Psychol.*, in press.

MURRAY, E. J., WILLIAMS, H. L., & LUBIN, A. Body temperature and psychological ratings during sleep deprivation. *J. exp. Psychol.*, 1958, **56**, 271–273.

MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

SEARS, R. R., HOVLAND, C. I., & MILLER, N. E. Minor studies of agression: I. Measurement of aggressive behavior. *J. Psychol.*, 1940, **9**, 277–281.

SPENCE, D. P. A new look at vigilance and defense. *J. abnorm. soc. Psychol.*, 1957, **54**, 103–108.

TYLER, D. B. The effects of amphetamine sulfate and some barbiturates on the fatigue produced by prolonged wakefulness. *Amer. J. Physiol.*, 1947, **150**, 253–262.

VEROFF, J. Development and validation of a projective measure of power motivation. *J. abnorm. soc. Psychol.*, 1957, **54**, 1–8.

WISPE, L. G. Physiological need, verbal frequency, and word association. *J. abnorm. soc. Psychol.*, 1954, **49**, 229–234.


Received May 2, 1958.

# SUGGESTIBILITY AS DEPENDENCY BEHAVIOR[1]

LEONARD F. JAKUBCZAK AND RICHARD H. WALTERS

*University of Toronto*

AN ANALYSIS of experimental procedures used in studies of suggestibility supports an interpretation of suggestibility as a form of dependency behavior. In the experiments of Sherif (1935) and Asch (1940), designed to demonstrate the influence of group norms, suggestions were given to subjects (*S*s) which resulted in increased conformity of their judgments to those of other members of the group. The changes in judgments were said by Asch to result from *S*'s need for social support, in other words, they were evidence of *S*'s dependency on the group. When the autokinetic effect is employed as a means of demonstrating the influence of suggestion, *S* is confronted with a highly ambiguous stimulus and is asked to make judgments in the absence of customary cues. A second person also makes judgments that *S* may accept as criteria against which to validate his own. Under these circumstances the judgments of another person become cues which *S* may accept or reject. If *S* accepts them, he is accepting help; if he accepts help, he is responding in a dependent manner.

Both experimental and clinical evidence support an interpretation of suggestibility in terms of dependency behavior. Inverse relationships between suggestibility and ascendancy (Kelman, 1950) and between suggestibility and self-sufficiency (Young & Gaier, 1953) have been found. Moreover, suggestibility and dependency have been traditionally associated in clinical descriptions of hysterical reactions. However, in spite of implicit and explict treatment of suggestibility in terms of dependency behavior, there has been little attempt to relate suggestibility to the development of dependency and independence in young children.

Sears, Whiting, and their collaborators (Sears, Maccoby, & Levin, 1957; Sears, Whiting, Nowlis, & Sears, 1953; Whiting & Child, 1953; Beller, 1955) have regarded dependency as a learned motivational system of which help-seeking and approval-seeking are manifestations. From this point of view, dependency behavior is acquired in the earliest relationship of the child to its parents, and is generalized to other adults and also, to a lesser extent, to peers. Dependent responses are necessarily to some degree reinforced in the first year or two of life, if only because the young child is too helpless and immature to care for its own needs. On the other hand, independence training also commences early and results in the partial inhibition of dependency behavior. Thus the conditions for a conflict in the area of dependency are present very early in the life of every child.

Miller (1948) has provided a theoretical model for the displacement of responses under conditions of conflict. This model assumes that the gradient of generalization of an inhibitory response is steeper than that of the response which it inhibits. The assumption permits predictions about the strength of overt responses, under conditions of strong and of weak inhibition, both to the persons in whose presence the conflict is originally learned and to persons of increasing dissimilarity from the original instigators of conflict. The relative strength of displaced overt responses may be depicted by subtracting the height of the curve representing the inhibitory response from the height of the curve representing the response that is inhibited, and plotting this set of differences as a new curve (Whiting & Child, 1953, p. 293). Since dependency on the parents is rarely, if ever, completely inhibited in children, only the right-hand portion of Miller's

[1] This paper is based in part on a thesis submitted by the first-named author in partial fulfilment of the requirements for the degree of M.A. at the University of Toronto. The authors wish to express appreciation to G. B. Thornton and E. Tulving for constructive advice and assistance and, also, to the Institute of Child Study of the University of Toronto and the Forest Hill Village Board of Education for their cooperation in this study. They wish in particular to express gratitude to Dorothee Keschner for permission to adapt and revise her test of dependency and independence and for her ready assistance in the development of a parallel test for older children.

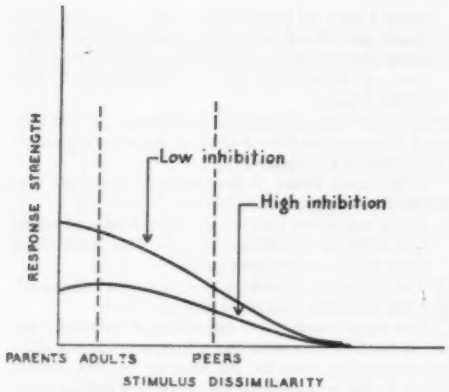


FIG. 1. HYPOTHETICAL STRENGTH OF DISPLACED DEPENDENCY RESPONSES AT VARYING POINTS ON A STIMULUS DISSIMILARITY CONTINUUM FOR CONDITIONS OF STRONG AND WEAK INHIBITION

model was taken into account in constructing curves for predictions concerning dependency behavior (Fig. 1).

Dependency behavior was regarded, for the purposes of this study, as a class of responses that included both help-seeking behavior and the acceptance of suggestions. High-dependent children were defined as those who readily accepted help even in tasks that they could perform by themselves; low-dependent children as those who showed reluctance to accept help even when a task was beyond their level of skill. In view of the description of the development of dependency and independence provided by Sears, Maccoby, and Levin (1957), it was assumed that the low-dependent group showed considerable inhibition of dependency responses, whereas the high-dependent group showed little such inhibition. On the basis of these assumptions, and those reflected in the hypothetical curves shown in Fig. 1, the following hypotheses were developed:

1. High-dependent children are more suggestible than low-dependent children.

2 Children respond more strongly when adults give suggestions than when peers give suggestions.

3. The difference in suggestibility between high-dependent and low-dependent children is greater when suggestions are given by adults than when suggestions are given by peers.

## METHOD

### *Subjects*

Subjects (*Ss*) were chosen on the basis of a modified form of a test of dependency and independence first developed by Keschner (1957) for use with five- to eight-year-old children.[2] For the purposes of this study, the test was extended for use with eight- to ten-year-old boys. The modified test consisted of 24 cartoon-like drawings, each depicting an adult offering help to a child. The child was asked to indicate, by circling Yes or No beneath each drawing, whether he would, or would not, accept help. The test was administered to 60 nine-year-old boys in an elementary public school. At the same time, a list of the situations depicted in the test was sent to the boys' parents who were asked to indicate whether or not the child could carry out the tasks unaided.

If a child indicated that he would accept help when his parent indicated that he could carry out the task in question, he was given a score of 1; if he indicated that he would not accept help when his parent indicated that he could not carry out the task in question, he was given a score of −1. The 12 boys with the highest total score for the 24 situations were selected as high-dependent *Ss*; the 12 boys who scored lowest were selected as low-dependent *Ss*.

Each *S* was twice exposed to suggestions, once by a peer and once by an adult. Half the high-dependent *Ss* and half the low-dependent *Ss* were exposed on the first experimental session to the peer and on the second session to the adult; the remaining *Ss* were exposed to the adult on the first session and to the peer on the second session. *Ss* were randomly assigned to the two orders.

### *Confederates*

Four eleven-year-old boys and four adults were trained to give suggestions. Three boys and three adults bore the major load of suggesting; the remaining two confederates were substitutes in case the scheduled suggester was not available at the time he was required. Each regular confederate had four high-dependent and four low-dependent *Ss* assigned to him. Care was taken to insure that the average time interval between the first and second exposure to suggestion was approximately the same for the high-dependent and the low-dependent *Ss*, and for the two orders of suggestion. The substitute adult and peer suggesters had each to be used on two occasions to keep time intervals controlled. First sessions were completed for all *Ss* before second sessions were begun; as far as other conditions allowed, high-dependent and low-dependent *Ss* were run in random order.

[2] A copy of the revision of Keschner's test for dependency and independence used in this study has been deposited with the American Documentation Institute. Order Document No. 5903 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.75 for microfilm or $2.50 for photocopies. Keschner's original test is included in her doctoral thesis (University of Toronto, 1957).

### *Apparatus*

The apparatus used for the autokinetic sessions was an approximate duplicate of that described by Sherif (1935) and used by more recent investigators (Kelman, 1950; Linton, 1954). It was located at one end of a photographic darkroom, 7 ft. from *S*, and was screened from *S*'s view when the autokinetic effect was not actually being induced.

Both *S* and the confederate were given toy "crickets" which they used to indicate that the light was beginning to move. The "cricket" used by the confederate was slightly muted to allow *E* to identify the producer of the signal. *S* was supplied with a pencil and pad on which to record his judgments concerning the apparent movement of light. In order to prevent *S* from suspecting that the confederate was not just another *S*, the latter also wrote on a pad at appropriate times during the experiment.

Crutchfield and Edwards (1945) found that the hum of a motor enhances the autokinetic effect. Consequently, to simulate the sound of a motor, a portable electric fan was located immediately behind the autokinetic apparatus and was kept in operation throughout the experiment.

### *Procedure*

*S* was given the following preliminary instruction immediately before entering the experimental room:

> I'm doing an experiment. I'm trying to find out how well people see a light move in a completely dark room. I want you to look at this light, and tell me how many inches it moves. O.K?

*S* was also told that another person was already in the room, and that he, too, was taking part in the experiment. When *S* was seated in a chair, adjacent to the suggester, *E* gave the following instructions:

> In a little while, I shall close the door and turn off the light, so that it will be completely dark in here. Then I shall turn on a little point of light. It will be located directly in front of you. I want you to look for this light, and as soon as you see it, let me know. Then we'll start the experiment. I'll turn on the motor, and then the light. I want you to look at it very carefully, and signal me as soon as you see it starting to move. Signal me by pressing your "cricket," like this [*E* demonstrates]. A few seconds after you have signaled, the light will be turned off. Then I want you to take your pad and pencil, and write down how many inches you think the light has moved. After you write down the number, turn the page over, so that there is only one number on each page. O.K.? Any questions?

*Pretraining Trials.* On each trial, the autokinetic light was turned on for a period of approximately 5 sec. beyond the time at which *S* signaled, or for a period of 20 sec. if no signal was made. The confederate signaled immediately after *S* signaled. This procedure was followed for eight trials.

*Training Trials.* *S* was now instructed as follows:

> Now in this part of the experiment, I want both of you to call out your judgments after writing them down. So for the next few trials, you do the same thing that you did before, except that you also call out how far you think that the light has moved. Since I have to keep track of who gives the judgment, why don't you answer in a definite order? Let's say you (*S*'s first name), and then you, Mr. —— (if adult), or you —— (confederate's first name, if peer).

After *S* gave his estimate, the confederate gave an estimate 5 in. greater than *S*'s. Eight training trials were carried out in this manner.

*Posttraining Trials.* A final series of four trials was given. *S* was instructed as follows:

> In the remaining part of this experiment, keep quiet and don't say anything out loud. Just mark down your judgments on your pads.

At this point, the experimental session was terminated, and the child was allowed to return to class.

The second session was carried out in the same manner as the first, except for two adjustments in procedure necessitated by some *S*s' reactions to the first session. Since, in the pretraining trials of the first session, some *S*s appeared to delay signaling until the confederate had indicated that he had seen the light move, during the second session the confederate signaled before *S* on every other pretraining trial. Two or three *S*s had noted that the confederate's judgment had always been 5 in. greater than theirs. Consequently, during the second session, instead of adding five to each of *S*'s estimates, the confederate added a variable amount, averaging five over eight trials. The implications of these changes are considered later in the discussion.

One *S* was lost from the low-dependent group on the second session. This *S* claimed that he could no longer see the light and, consequently, further training was impossible.

### *Measures*

A suggestion that the light would move was given at the outset of the experiment. This suggestion was reinforced during the pretraining trials by the confederate's signal indicating that he had seen the light move. Consequently, one measure of suggestibility could be obtained from *S*'s responses in the pretraining trials. The median of *S*'s responses on the pretraining trials was taken as his pretraining response. The effect of training was measured by subtracting *S*'s median response on the pretraining trials from his median response on the posttraining trials.

Not every *S* wrote down a response each time he signaled with his "cricket." In addition, some *S*s gave an occasional extreme and unrealistic response. A nonparametric analysis of data was clearly indicated.

## RESULTS

Table 1 shows the median pretraining judgment of each *S* and, also, the difference between his median pretraining and his median posttraining judgment for both sessions.

On the pretraining trials of the first session, high-dependent *S*s were significantly more suggestible than were low-dependent *S*s ($U = 41$; $p < .05$). A finer analysis of results on these trials, which took into account the status of the suggester, revealed a marked difference

TABLE 1
MEDIAN RESPONSES OF HIGH-DEPENDENT AND LOW-DEPENDENT SUBJECTS TO ADULTS AND PEERS ON FIRST AND SECOND SESSIONS[a]

| Order | Subject | First Session | | Second Session | |
|---|---|---|---|---|---|
| | | Pretraining Trials | Effect of Training[b] | Pretraining Trials | Effect of Training |
| High-dependent *Ss* | | | | | |
| Peer first | 9 | 1.50 | 38.50 | 7.00 | 36.50 |
| | 11 | 1.00 | 5.00 | 2.00 | 4.00 |
| | 8 | 2.50 | 11.00 | 4.00 | 6.50 |
| | 10 | 1.00 | 3.00 | 6.50 | 2.50 |
| | 7 | 3.00 | 49.50 | 22.00 | 68.00 |
| | 12 | 2.00 | 4.00 | 2.00 | 2.00 |
| Adult first | 5 | 3.50 | −1.00 | 3.00 | −0.50 |
| | 3 | 3.00 | 11.00 | 9.00 | 2.00 |
| | 2 | 2.50 | 18.00 | 18.00 | −3.00 |
| | 1 | 3.00 | 3.00 | 3.00 | 9.50 |
| | 25 | 4.00 | 2.50 | 9.00 | 0.00 |
| | 26 | 3.50 | 24.00 | 15.00 | 31.00 |
| Low-dependent *Ss* | | | | | |
| Peer first | 20 | 1.00 | 19.50 | 13.00 | 27.00 |
| | 21 | 0.00 | 12.50 | 13.00 | 3.50 |
| | 23 | 0.00 | 4.50 | 6.00 | 7.50 |
| | 22 | 2.00 | 9.50 | 5.00 | 0.50 |
| | 24 | 2.50 | 6.00 | 4.00 | 6.50 |
| | 19 | 2.00 | 55.50 | 12.00 | 638.00 |
| Adult first | 17 | 2.50 | 9.00 | 5.00 | 2.00 |
| | 18 | 2.00 | 3.50 | 2.00 | 1.50 |
| | 16 | 2.00 | 5.50 | 4.50 | 2.50 |
| | 13 | 2.50 | 7.00 | 4.00 | 3.00 |
| | 14 | 3.00 | 14.50 | 14.00 | 6.00 |
| | 15 | 2.00 | 6.00[c] | | |

[a] Responses reported in terms of inches.
[b] Calculated by subtracting *S*'s median pretraining response from his median posttraining response.
[c] This *S* failed to make any response on the second session.

between high-dependent and low-dependent *Ss* in their suggestibility to adults ($U = 3$; $p = .008$); on the other hand, the difference between high-dependent and low-dependent *Ss* in their suggestibility to peers was much smaller and failed to reach significance. No significant difference between high-dependent and low-dependent *Ss* was found after the first pretraining session.

Data for high-dependent and low-dependent *Ss* were now combined for assessing differences attributable to the status of the suggester. On the pretraining trials of the first session adults were significantly more effective in their suggestions than were peers ($U = 21$; $p < .01$). In addition, adults were more effective than peers in training during the second session ($U = 32.5$; $p < .01$). There was no difference in the effectiveness of adults and peers in training during the first session or on pretraining trials of the second session.

The order of exposure to adult and peer suggesters had an important influence on results. Both the adult and the peer were significantly more effective in the second, than in the first, pretraining session ($U = 22$; $p < .01$ for adults; $U = 5.5$; $p < .001$ for peers). The adult was no more effective in changing responses from the pretraining to the posttraining sessions when he preceded the peer than when he followed the peer. On the other hand, the peer was much more effective in changing responses if he preceded the adult than if he followed him ($U = 20$; $p < .01$).

The greater effectiveness of both the adult and of the peer in the second pretraining session strongly suggested that transfer of learning from the first to the second session had occurred. Accordingly, transfer effect from the first to the second session, both for pretraining responses and for effects of training, was assessed by the Wilcoxon test (Siegel, 1956). While there was a highly significant increase in pretraining responses on the second session ($T = 1$; $p < .001$), the effect of training was, if anything, somewhat reduced.

In order to estimate both the effects of learning and of procedural changes during the experiment, trial-by-trial graphs of *Ss*' responses were constructed for each of the experimental sessions (Figs. 2 and 3). In order to give each *S*'s series of responses equal weight in determining the form of the curves, a nonparametric procedure was used in their construction. Each *S* should have made 20 responses on each experimental session, eight in the pretraining period, eight in the training period, and four in the posttraining period.

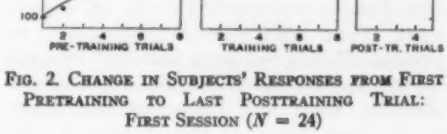


FIG. 2. CHANGE IN SUBJECTS' RESPONSES FROM FIRST PRETRAINING TO LAST POSTTRAINING TRIAL: FIRST SESSION ($N = 24$)

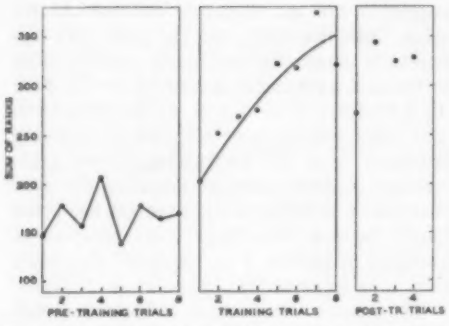


FIG. 3. CHANGE IN SUBJECTS' RESPONSES FROM FIRST PRETRAINING TO LAST POSTTRAINING TRIAL: SECOND SESSION ($N = 23$)

If *S* had not recorded a response, his median response for the series of four or eight trials, as the case might be, was substituted. Since *E* could not know on which trial *S* had failed to respond, substitute scores were randomly assigned to trials. A complete set of 20 scores was thus obtained for each *S*. These series of 20 responses were then ranked for each *S*. *S*'s smallest response during the series was given a rank of 1, his largest response a rank of 20. Twenty-four sets of ranked responses were thus obtained for the first session, 23 for the second session. The sum of the ranks for each trial was calculated and used to construct the graphs shown in Figs. 2 and 3.

## DISCUSSION

While predictions were not consistently confirmed, the results of the experiment provide considerable support for the interpretation of suggestibility in terms of dependency behavior. On the pretraining trials of the first session, all three experimental hypotheses were confirmed. It should be noted that in these pretraining trials the only suggestion given was the confederate's signal that he had seen the light move; no verbalized judgment was given.

Failure to obtain consistent differences beyond the first pretraining trials may be due in part to undetermined effects of the experimental procedure itself. Predictions were, in fact, based on hypothesized differences in child-training procedures prior to the experiment. These differences were expected to produce different responses from high-dependent and low-dependent *S*s in the experimental situation. While such differences were found at the very outset of the experiment, the experimental procedure may, by training all *S*s to respond in a dependent manner, have served largely to reduce differences which previously existed.

*S*s' responses were greatly influenced by the order of exposure to adult and peer suggesters. The increased responsiveness to both adult and peer during the pretraining trials of the second session can be simply explained as a positive transfer effect. Such an effect cannot, however, explain the finding that peers are less effective in producing change from pretraining to posttraining trials when they follow adults than when they precede them. If considerable learning occurs during the first session, possibly only an adult (presumably a stronger reinforcer than a peer) can produce a further increase in suggestibility.

The graphical presentation of data (Figs. 2 and 3) suggests that learning may occur even during the pretraining trials of the first session, though it is, of course, possible that the changes during these trials reflect a "warm-up" effect. These changes have, in any case, marked implications for prior studies of the autokinetic effect in which practice sessions have been given *S*s before the experiment proper begins. With child *S*s it is apparently unnecessary to make a judgment contrary to *S*s' in order to enhance the autokinetic effect; one has only to imply that the light has moved. It is possible that a similar effect could be produced with adults.

During the first session's training trials *S*s' responses form a negatively accelerated curve (Fig. 2) analogous to many learning curves. On the other hand, the curve for posttraining trials indicates that extinction begins when verbal suggestions are no longer given. Suggestion can thus be interpreted as reinforcement of dependency behavior.

The results for the second session are of considerable interest. During the pretraining trials of this session the confederate signaled before *S*s on half the trials and after *S*s on the other half. The effect of this alternation is clearly reflected in Fig. 3. *S*s' judgments are greater if they occur on a trial on which the confederate signals first and follow a trial on which their own signal is reinforced by a subsequent signal from the confederate, i.e., on occasions when their responses are doubly reinforced.

The addition of a variable increment, aver-

aging 5 in., to *Ss*' judgments in the training trials has approximately the same effect as the addition of a constant 5 in. on each trial. A rather steep, negatively accelerated curve is obtained both from *Ss*' responses in the first training period and from their responses in the second training period. On the second set of posttraining trials *Ss*' responses again decrease, but not in as steady a manner as on the first posttraining trials. The last three of these trials, however, again suggest that over a longer series of trials, gradual extinction might take place.

## SUMMARY

Two groups, each consisting of 12 nine-year-old boys, were exposed to the autokinetic effect. The low-dependency groups consisted of boys who indicated unwillingness to accept help in tasks which they were unable to accomplish themselves; the high-dependency group consisted of boys who indicated willingness to accept help even when they required none. Each child was twice exposed to the autokinetic effect; on one occasion, judgments contrary to the child's were given by an adult; on the other occasion, contrary judgments were given by a peer. Half the *Ss* in each group received suggestions first from an adult, then from the peer; for the other half, the suggesters were employed in the reverse order.

Predictions were made (*a*) on the assumption that suggestibility was a form of dependency behavior and (*b*) were partly based on Miller's (1948) theoretical model for the displacement of responses under conditions of strong and weak inhibition. These predictions were confirmed by the results for the pretraining trials of *Ss*' first experimental session. High-dependent *Ss* were significantly more suggestible than low-dependent *Ss*, and adults were more effective as confederates than were peers. In addition, the difference in suggestibility between high-dependent and low-dependent *Ss* was much greater when adults were used as confederates than when suggestions were given by peers.

Though findings were not consistent for the remaining stages of the experiment, confirmatory evidence of the greater effectiveness of adult suggesters was provided on the second session; on this session, adults produced a greater change in *Ss*' responses from pretraining to posttraining trials than did peers. The effect of order of exposure to adult and peer suggesters proved to be important, and complex transfer effects apparently occurred.

A graphical analysis of changes over the series of trials suggested that learning may occur even before contrary judgments are given by confederates in the training trials. Learning is considerably increased, however, when contrary judgments are given; steeply rising, negatively accelerated curves were obtained for the training sessions. When suggestion is no longer given in the posttraining trials, *Ss*' responses apparently begin to extinguish.

The findings as a whole support the interpretation of suggestibility in terms of dependency behavior, especially if dependency is itself regarded as a class of learned responses which may be reinforced or extinguished by the responses of other persons.

## REFERENCES

ASCH, S. E. Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards. *J. soc. Psychol.*, 1940, **12**, 433–465.

BELLER, E. K. Dependency and independency in young children. *J. genet. Psychol.*, 1955, **87**, 25–35.

CRUTCHFIELD, R. S., & EDWARDS, W. The effect of a fixated figure on autokinetic movement. *J. exp. Psychol.*, 1949, **39**, 561–568.

KELMAN, H. C. Effects of success and failure on suggestibility in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.

KESCHNER, DOROTHEE A. Dependence and independence in primary-school children. Unpublished doctoral thesis, Univer. of Toronto, 1957.

LINTON, HARRIET B. Rorschach correlates of response to suggestion. *J. abnorm. soc. Psychol.*, 1954, **49**, 75–83.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child training*. Evanston: Row Peterson, 1957.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–235.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol., N. Y.*, 1935, No. 187.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

YOUNG, M., & GAIER, E. L. A preliminary investigation into the prediction of suggestibility from selected personality variables. *J. soc. Psychol.*, 1953, **37**, 53–60.

WHITING, J. W. M., & CHILD, I. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

Received May 14, 1958.

# PERCEPTUAL PREFERENCES AS A FUNCTION OF PERSONALITY VARIABLES UNDER NORMAL AND STRESSFUL CONDITIONS[1]

NATALIA POTANIN
*Johns Hopkins University*

THE Rorschach technique rests on the assumption that the way a person perceives an inkblot means something about his personality. Although Rorschach theorists have suggested many hypotheses about the connection between personality variables and perceptual variables, difficulties are often encountered in establishing precise relationships between personality attributes and Rorschach scores. These difficulties arise from the following sources: The determinants used to obtain Rorschach scores have not been analyzed sufficiently to pinpoint the perceptual variables employed in the responses; the personality variables have not been adequately defined; and the relationships between personality variables and perceptual variables have not been tested in experimental settings. The present study explores the relationship between certain perceptual variables implicit in the Rorschach shading responses and the personality variables which are supposed to be related to them.

In the clinical situation the stimulus material, the inkblot, remains the same; it is the response of the person that varies. One way of explaining what a person does when he interprets an inkblot is to say that he selects certain features to respond to. Presumably, then, personality variables determine what features the person selects or "prefers" in the whole stimulus complex. If this is so, he should *prefer* the same features when he sees them in an objectively variable series of stimuli. Moreover, there should be some relationship between his preferences and certain personality characteristics. This rationale underlies the present investigation.

To apply this argument to the shading responses, we must first analyze what features are being responded to in the inkblot by analyzing the responses that are made to it. According to Klopfer (1954), whose scoring system provides the most detailed analysis of shading responses, one of the ways in which shading may be used is to give the impression of surface or texture. What perceptual variables underlie a "texture" response? When we say that something is textured, we generally mean that the surface contains patterned *detail* and varies in *depth*. Thus, a texture shading response may involve two perceptual variables, detail and depth. For this reason, the stimuli in this study were a series of simple geometric forms which differed in amount of detail and in degree of depth.

As for personality variables, the Rorschach literature contains numerous hypotheses about the shading responses. Klopfer (1954) suggests that texture responses are indicative of the "degree of awareness and differentiation of the person's needs for affection and dependency" (p. 271) and that shading responses other than texture are indicative of anxiety about these affectional needs (pp. 268–269). Piotrowski (1939) argues that "the shading responses are indicators of fear and anxiety" (p. 253). Beck (1945) associates the use of "heavy shadings" with a "chronic readiness to be upset" (p. 41). Eichler (1951) found that induced stress led to an increase in the weighted sum of the shading responses on the Rorschach. To sum up, shading responses have been variously associated with the need for affection and dependency, anxiety about affectional need, general anxiety, and induced stress. The present investigation used a measure of dependency and two measures of anxiety, insecurity and symptomatic anxiety. These were related to preferences for detail and depth under normal and stressful conditions.

All of these considerations can be formalized in the following set of hypotheses tested in this study:

1. Preference for detail is related to acknowledged dependency.

[1] This report is based on a dissertation submitted to Johns Hopkins University in partial fulfillment of the requirements for the degree of doctor of philosophy. The writer wishes to thank Mary D. Ainsworth and W. R. Garner for the many helpful criticisms and suggestions in carrying out the investigation. I am also indebted to A. Chapanis whose patient guidance improved the clarity of this paper.

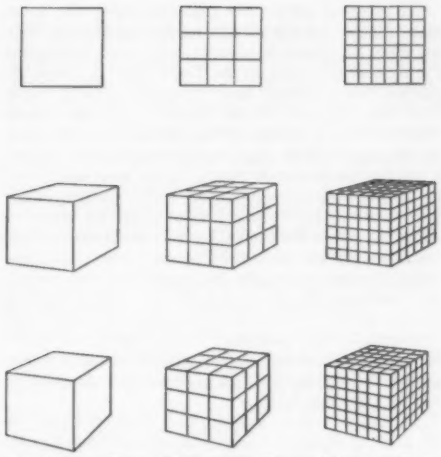

FIG. 1. THE NINE GEOMETRIC FORMS USED IN THIS STUDY. THE FORMS INCREASE IN DETAIL FROM LEFT TO RIGHT, AND IN DEPTH FROM TOP TO BOTTOM

2. Preference for detail is related to acknowledged anxiety.
3. Preference for depth is related to acknowledged dependency.
4. Preference for depth is related to acknowledged anxiety.
5. Preference for detail changes under experimentally induced stress.
6. Preference for depth changes under experimentally induced stress.

## METHOD

The visual stimuli consisted of nine geometric forms which varied in depth and detail (see Fig. 1). The nine forms were paired in all possible combinations (72, including both positions), and eight basic orders of 36 paired comparisons were generated by using the Ross (1939) matrix for minimizing time and space errors. The pairs of forms were projected on a 4′ x 4′ screen to produce an image of maximum size and exposed for exactly three seconds. *Ss* were seated about 7 ft. away from the screen. They were asked to indicate on a data sheet which of each pair of forms they liked better.

### *Personality Variables*

*Criterion of dependency.* An index of dependency was obtained by summing the immature dependent and mature dependent subtest scores on the Ainsworth (1958) tests of extra-familial and avocational security-insecurity. These items measure the extent of an individual's awareness of dependency relationships with others. In immature dependency, the individual is aware of being the passive recipient of need satisfactions through others. In mature dependency, the individual realizes that his satisfaction rests upon mutually contributing relationships and tries to reciprocate by interpreting, anticipating, and fulfilling the needs of others. The sum of the scores on these two scales was taken to be the measure of dependency.

*Criteria of anxiety.* Two related tests of anxiety were used, one measuring feelings of insecurity, the other, symptomatic anxiety. The index of insecurity was the summed scores on the insecurity subtests of the Ainsworth (1958) tests of familial, extra-familial, and avocational security-insecurity. This measure reflects the extent to which a person admits to such feelings as loneliness, isolation, boredom, and tension as a result of a failure to find satisfaction of his need for intimate relationships with other people. The Taylor Manifest Anxiety Scale (1953) was used as the criterion of symptomatic anxiety. It is an inventory of symptoms and feelings of anxiety that an individual acknowledges without committing himself about the specific problems causing his anxiety.

The insecurity and dependency subtests of the Ainsworth tests have nonoverlapping items. In addition, the dependency scores were unrelated to the anxiety scores, measured either as insecurity ($r = +.08$) or as symptomatic anxiety ($r = -.10$). The two measures of anxiety, however, were highly correlated ($r = +.66$). Thus the extent to which an individual admits his dependency on others seems unrelated to his feelings of insecurity or general anxiety.

### *Induced Stress*

Because male college adolescents are particularly sensitive about their masculinity, a stressful situation was successfully created by telling the *Ss* that they were on the "borderline" according to scores obtained on a masculinity–femininity (M-F) questionnaire. All *Ss* were later convincingly reassured about their test scores. The very personal nature of the stressful situation discouraged any leakage of information about the stress situation.

### *Subjects*

The *Ss* were 148 male undergraduates enrolled in the introductory psychology class at Johns Hopkins University. They come from a student population which has a median IQ of about 123 and is composed of people mainly from the upper-middle socioeconomic class. The *Ss* were mostly freshmen about 18 years old. No student who was over 24 years or who was known to be married was included in the sample.

*Ss* were selected from a total of 185 students on the basis of their scores on the dependency and insecurity scales administered to the whole class prior to any experimentation. Two overlapping subsamples of *Ss* were selected so as to form three balanced groups on each of these two scales. The following 100 *Ss* were selected on the basis of their dependency scores:

1. Low dependency group: 24 *Ss* from the lowest quarter of the distribution of scores, range 3–11.
2. Moderate dependency group: 52 *Ss* from the middle half, range 12–17.
3. High dependency group: 24 *Ss* from the highest quarter, range 18–22.

The following 132 *Ss* were selected on the basis of their insecurity scores:

1. Low insecurity group: 44 *Ss* from the lowest quarter of the distribution of scores, range 0–6.
2. Moderate insecurity group: 44 *Ss* from the middle half, range 7–14.
3. High insecurity group: 44 *Ss* from the highest quarter, range 15–28.

These two groups together comprise the 148 *Ss* used in this investigation. For the purposes of analysis, a subsample of 100 *Ss* was selected from this total on the basis of their scores on the Taylor anxiety scale:

1. Low anxiety group: 27 *Ss* from the lowest quarter of the distribution of scores, range 0–8.
2. Moderate anxiety group: 52 *Ss* from the middle half, range 9–18.
3. High anxiety group: 27 *Ss* from the highest quarter, range 19–46.

### *Procedure*

The dependent variable for preferences under normal conditions was the number of times that each geometric form was preferred. *Ss* attended the sessions in small groups of two to six people, and preference for the forms was established under the guise of research on a grant from an interior decorating firm. *Ss* were asked not to think about the forms but to react to each pair emotionally. The preferences could be summed in two ways to give either a measure of preference for detail irrespective of depth, or a measure of preference for depth irrespective of detail. The dependent variable in preference under stressful conditions was the difference between the prestress and poststress preferences in terms of either detail or depth.

The experiment was conducted over a period of two sessions exactly a week apart. *Ss* were asked about their preferences for the geometric forms on both occasions. Half of the *Ss*, the experimental group, were exposed to stress at the second session, while the other half, the control group, were not exposed to stress. Each of these groups was further subdivided so as to form two stress groups, $E_1$ and $E_2$, and two nonstress groups, $C_1$ and $C_2$. These four groups were matched in terms of insecurity scores in such a way that the means and standard deviations at each level of insecurity were alike. In addition, the two stress groups ($E_1$ and $E_2$) were further matched in terms of mean dependency scores at each level of insecurity.

During the first session, initial preferences for the geometric forms were established for Groups $E_1$, $C_1$, and $C_2$, but not for Group $E_2$. The *Ss* in Group $C_2$ were then dismissed. The *Ss* in Groups $E_1$, $E_2$, and $C_1$ were led to believe that the personality tests which followed were part of an entirely different investigation. The *Ss* completed an abridged version of the Masculinity-Femininity (M-F) scale of the MMPI. While some of these were supposedly being scored by the experimenter, they answered the Taylor anxiety questionnaire. The experimenter reassured the nonstress *Ss* ($C_1$) about their "masculinity" by calling them aside and telling them confidentially that they had scored above average in the "moderately high masculine" category. The stress group *Ss* ($E_1$ and $E_2$) were told to expect their results on the M-F test the following week.

At the next session the following week, the stress group *Ss* ($E_1$ and $E_2$) were pointedly called away from the other *Ss*, placed in separate rooms, and told rather sternly that they were on the "borderline" in the previous M-F test and would have to complete a much longer M-F test (the rest of the MMPI items, plus others devised for the occasion). While waiting for their results on the second M-F test, the *Ss* repeated their preferences for the geometric forms. These were considered to be the preferences obtained under experimentally induced stress. Each *S* in the stress group was reassured by calling him aside and telling him confidentially that the score on his second M-F test placed him very definitely above average in the "moderately high masculine" category. It was explained that "borderline" in his case meant on the border between "very high" and "moderately high masculine." The *S* was left with the feeling that it had been foolish of him to have misinterpreted "borderline." All *Ss* accepted this explanation readily and thankfully.

During the second session, the nonstress Group *Ss* ($C_1$ and $C_2$) repeated their preferences for the geometric forms, and the *Ss* in Group $C_1$ departed. The *Ss* in Group $C_2$ completed the Taylor anxiety scale before leaving.

Thus, Group $E_1$ may be termed the standard experimental stress group and Group $C_1$ the standard nonstress control group. Group $E_2$ controlled for any type of "learning" or stereotypy that might result from a repetition of the perceptual preferences; these *Ss* followed exactly the same procedure as those of Group $E_1$, except that the prestress preferences were omitted. In order to make sure that preference for the geometric forms was not influenced in any way by the introduction of the personality tests, Group $C_2$ *Ss* did not complete the Taylor anxiety scale until the end of the second session. This arrangement of the groups made it possible to derive two estimates of change in preference under stress: (*a*) the difference in preference scores between the first and second sessions for Group $E_1$ as compared to the difference in scores for the combined Groups $C_1$ and $C_2$, and (*b*) a comparison of the initial prestress preferences of Group $E_1$ with the poststress preferences of the matched Group $E_2$.

## RESULTS

Analysis of the results shows a substantial relationship between preference for detail and degree of acknowledged dependency. The preference for detail by the three dependency groups was significantly different at better than the .001 level of significance with either a chi square test or an analysis of variance. The linear correlation coefficient ($r$) of $+0.29$ between dependency score and number of preferences for the highly detailed forms was also significant ($p = .01$). Preference for detail was apparently not related to either insecurity or Taylor anxiety scores. Thus the first hypothesis was strongly upheld, but the second hypothesis was not substantiated.

There was no evidence, however, of any relationship between preference for depth and any of the three personality measures used in this investigation. All *S*s preferred the cube of moderate depth the most, and the square the least. Thus, the third and fourth hypotheses were not substantiated.

There were only slight changes in preference for detail under stressful conditions as compared to normal conditions, but those changes which did occur were a function of the insecurity scores. As indicated previously, there were two ways of measuring change in preference under stress: first, by taking the difference between each *S*'s preferences under normal and stressful conditions and comparing the mean differences for the *S*s in Group $E_1$ with the mean differences of *S*s in Group $C_1$ and $C_2$ combined; and, second, by taking the mean difference between the initial prestress preferences of Group $E_1$ and the initial poststress preferences of the matched Group $E_2$.

With the first method of measuring the effects of stress, an analysis of variance showed the interaction *Insecurity* × *Stress* as significant at the .05 level. Under stress, low insecurity *S*s increased their preference for detail; moderate insecurity *S*s decreased their preference for detail, and the high insecurity *S*s showed almost no change. The nonstress control *S*s ($C_1$ and $C_2$) showed only slight changes in preference, but, here again, the high insecurity *S*s showed the least change.

The second method of measuring the effects of stress did not yield statistically significant results, but the trend of the differences is similar to that shown by the first method of analysis. In this case, however, the high insecurity *S*s showed the greatest decrease in preference for detail. Since high anxiety *S*s tend to show perceptual perseveration (Moffitt & Stagner, 1956), the lack of change exhibited by the high insecurity *S*s in the first method of analysis may be a result of this perceptual rigidity.

Since change in preference for detail was unrelated to either dependency scores or Taylor anxiety scores, the fifth hypothesis may be regarded substantiated only in part.

Under conditions of stress, there was a small increase in preference for the two-dimensional forms, the squares. However, this increase was due entirely to the individuals who obtained a low score on the Taylor anxiety scale (see Table 1). An analysis of variance indicated a significant interaction of *Anxiety* × *Stress* at the .05 level. Change in preference for depth was unrelated to either dependency scores or insecurity scores. Thus, the sixth hypothesis was substantiated in part.

TABLE 1

MEAN CHANGE IN PREFERENCE FOR TWO-DIMENSIONAL FORMS UNDER STRESSFUL CONDITIONS

(A plus score indicates an increase in preference)

| | Stress Group $E_1$ ($N$ = 38) | Nonstress Groups $C_1$ and $C_2$ ($N$ = 68) |
|---|---|---|
| Low Anxiety group | +3.00 | −0.05 |
| Moderate and High Anxiety groups | +0.39 | +0.29 |

Because the finding that some *S*s prefer detail might be interpreted as a manifestation of obsessive-compulsive trends, the data were reanalyzed in terms of Psychasthenia (*Pt*) scores of the MMPI, which were available for most of the *S*s from another experiment. Neither initial preferences nor changes in preference under induced stress were related to *Pt* scores. Furthermore, the correlation between dependency scores and *Pt* scores was zero (−.04). This suggests that the kind of detail preferred in this investigation is not the isolated, partitioned detail that the obsessive-compulsive uses.

Since dependency was related to detail, there was the possibility that *S*s who are considered clinically to be nondependent prefer the nondetailed forms. However, the Psychopathic Deviate (*Pd*) scores of the MMPI were unrelated to either initial preferences or changes in preference. The correlation between *Pd* scores and dependency scores was zero (−.04). Whatever the dependency scale is measuring, it is not the inverse of the *Pd* scale.

## DISCUSSION

The results of this study suggest that initial perceptual preferences are determined by one personality variable, but that changes in these preferences under induced stress may be determined by another quite independent personality variable or variables. A person's preference for detail is related to the amount of

dependency he acknowledges, and it seems independent of his level of insecurity and symptomatic anxiety. On the other hand, the change in his preferences under induced stress appears unrelated to his initial preferences and level of dependency, but is a function of his level of insecurity and general anxiety.

To interpret properly the relationship between dependency and preference for detail, we must recognize that the kind of "detail" used in this study is actually patterned or organized detail. Moreover, the lack of correlation between preference for detail and obsessive-compulsive trends shows that it is not *isolated* detail that is relevant in this study. Taken together, these facts suggest that dependency is related to preference for detail which forms part of a complex whole, rather than for isolated detail.

The relationship between dependency and preference for detail may be explained as follows: An individual who acknowledges his dependency on others must be aware that the behavior of others affects his well-being. To be so aware, he must be sensitive to nuances in the behavior of others. If this sensitivity is accompanied by the satisfaction of his needs by others, the individual attends to these nuances or details. If the person does not acknowledge his dependency on others for the satisfaction of his needs, he lacks motivation to attend to or prefer such nuances. The extent to which a person habitually attends to details when observing the behavior of others may generalize to other situations. If so, it would account for the observed relationship between dependency and preference for detail.

Since preference for detail is related to dependency, we may interpret the change in preference for detail under stressful conditions as a change in the extent to which dependency is acknowledged. The finding that low insecurity *S*s increase their preference for detail under stress seems to indicate that only secure individuals are able to acknowledge a stronger awareness of the need for others when placed in a stressful situation. *S*s who are more insecure decrease their preference for detail under stress, probably because they defend themselves against the stress by decreasing their awareness of their need for others.

The increase in preference for two-dimensional forms under stress resembles what *S* does when he makes a *k* response on the Rorschach. In a *k* response, *S* projects a three-dimensional expanse onto a two-dimensional plane. In this experiment, *S*s changed their preferences for three-dimensional forms to two-dimensional representations. Klopfer's (1954) hypothesis is that a *k* response indicates "affectional anxiety behind a good front of outward control and is found with subjects who cover up their anxiety with an intellectual cloak" (p. 279). Extending this interpretation to the present investigation, an increase in preference for two-dimensionality represents an increase in affectional anxiety which is covered up intellectually. The fact that it was only low anxiety *S*s who increased their preference for two-dimensionality suggests that it is only the generally nonanxious person who is able to rally intellectual defenses under conditions of stress.

The results of this study have shown that neither insecurity nor symptomatic anxiety is related to initial preferences for detail or depth. There is still the possibility that other perceptual characteristics of the blot material, such as degree of saturation of achromatic color, may be associated with these anxiety states as Piotrowski (1957) suggests. All that this investigation has shown is that those aspects of the stimulus material which are equivalent to "detail" as defined here, are related to acknowledged dependency but not to anxiety.

## SUMMARY

The Rorschach literature hypothesizes that shading responses are related to various personality variables such as needs for affection and dependency, anxiety about affectional needs, chronic anxiety, and induced stress. This investigation has studied the relationship between preference for detail and depth, which are assumed to underlie "texture" responses, and measures of dependency, insecurity, and anxiety under normal and stressful conditions. The results were as follows:

1. The more an individual acknowledges his dependency on others, the greater is his preference for detail.

2. Under conditions of induced stress, the change in preference for detail is a function of the level of the individual's insecurity; low insecurity *S*s increase their preference for

detail, but moderate and high insecurity *S*s decrease or do not change their preference for detail.

3. Under conditions of induced stress, the change in preference for depth is a function of the level of the individual's symptomatic anxiety; low anxiety *S*s increase their preference for two-dimensionality, but moderate and high anxiety *S*s do not change their preference for depth.

An interpretation of the relationship between preference for detail and level of dependency given here is that the individual generalizes from his sensitivity to nuances in the behavior of other people. The change in preference for detail with induced stress was interpreted as a change in awareness of dependency. The change in preference for depth with induced stress was interpreted as an intellectual type of defense against anxiety.

## REFERENCES

AINSWORTH, MARY D., & AINSWORTH L. H. *Measuring security in personal adjustment.* Toronto, Canada: Univer. Toronto Press, 1958.

BECK, S. J. *Rorschach's test: A variety of personality pictures.* Vol. II. N. Y.: Grune & Stratton, 1945.

EICHLER, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344–355.

KLOPFER, B., AINSWORTH, MARY D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique: Technique and theory.* Vol. I. Yonkers-on-Hudson, N. Y.: World Book, 1954.

MOFFITT, J. W., & STAGNER, R. Perceptual rigidity and closure as a function of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354–357.

PIOTROWSKI, Z. A. *Perceptanalysis.* N. Y.: Macmillan, 1957.

ROSS, R. T. Optimal orders in the method of paired comparisons. *J. exp. Psychol.*, 1939, **25**, 414–424.

TAYLOR, JANET A. A personality test of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Received May 14, 1958.

# VERBAL DISCRIMINATION OF PLEASANT AND UNPLEASANT STIMULI PRIOR TO SPECIFIC IDENTIFICATION[1]

CHARLES W. ERIKSEN, HIROSHI AZUMA, AND ROSALIE B. HICKS

*University of Illinois*

UNCONSCIOUS discrimination is not only a controversial issue in current psychology, but even among those who espouse such a view there is considerable confusion as to just how much discrimination can occur at unconscious levels and how precise such discrimination can be. Blum (1955) as well as Lazarus and McCleary (1951) imply an unconscious discriminating process which is not only temporally prior to conscious discrimination but which is also precise enough to discriminate a word like "shil" from a structurally quite similar but vulgar Anglo-Saxon word. Such a conception of a super-discriminating unconscious does not seem to be justified by a survey of the experimental literature (Eriksen, 1958), but the possibility would seem to exist that rather gross discriminations can occur at levels not accessible to conscious awareness.

Lazarus (1956), in discussing the subception effect, has suggested that gross dichotomous discriminations between threatening and nonthreatening stimuli might be made at levels below awareness, and Osgood (1957) has given a more explicit statement of this possibility. Osgood (1957) conceives that an over-all affective evaluation occurs early in the perceptual process and "that hints provided by the incomplete information in stimulus displays, eliciting the dominant evaluative (emotional) components of mediating reactions which feed cues back into the perceptual system, may produce either facilitation of co-valuent integrations (vigilance) or disorganization of integrative mechanisms (defense)" (p. 307).

These suggestions not only seem plausible, but upon cursory examination seem to escape the need for ascribing to the unconscious as precise a discrimination as that which occurs in awareness. Binder (1955), among others, has pointed out that one or more characteristics of a stimulus may be discriminated in the absence of sufficient discrimination of the remaining characteristics to permit specific identification. Thus in a brief exposure we may be able to discriminate the color of a stimulus without perceiving its texture or shape. And if we can discriminate the color of a stimulus without also perceiving its texture and shape, why could we not be expected to discriminate the affective value of a stimulus in the absence of specific identification?

The difficulty with such a conclusion is this. Generally, the meaning of a stimulus is conveyed not by one or two physical characteristics of the stimulus but by the total aggregate of characteristics. In the case of words, which are frequently used as stimuli in experiments in this area, the affective meaning is conveyed only by the complete word, not by isolated characteristics such as word length, or that there is a tall letter at the beginning of the word, etc. For discrimination of affective meaning to occur, somewhere in the nervous system the complete word would have to be identified. Thus while Osgood and Lazarus may be correct in their views that affective discrimination occurs temporally prior to conscious discrimination, it must be recognized that this requires a very precise discrimination at unconscious levels.

There is an alternative theory that does not require assumptions of either prior or more precise unconscious discrimination but accounts for correct discriminative responses in the absence of conscious verbal discrimination. Elsewhere the senior author (Eriksen: 1956a, 1956b, 1958) has shown that subception and similar effects can be accounted for in terms of concurrent response systems that have independent sources of error. Perception is assumed to give rise to a number of different responses of which verbal identification is only one. Exposure of an emotion-laden stimulus not only leads to verbal identification but concurrently activates affective responses. If the exposure of the stimulus is very brief or otherwise impoverished, the perceptual process is weak,

[1] This research was supported by Research Grant M-1206 from the National Institute of Mental Health, Public Health Service.

and there is considerable error in both the verbal identification and the affective response systems. However, if the errors in these two response systems are less than perfectly correlated and both systems are operating at above chance accuracy, then the affective responses would show better than chance discrimination on those trials where the verbal identifications were incorrect and vice versa. In terms of this formulation, affective responses are not assumed to occur temporally prior to conscious verbal identification, but concurrently. Also, the threshold for such affective discrimination (which traditionally has been equated with unconscious discrimination) does not have to be lower than that for verbal identification but can actually be the same or even higher. The subception-like effect will occur as long as both the affective response, and the verbal identification are responding to the perceptual process and have at least partially independent sources of error.

In the experiment reported below we have investigated the capacity of human subjects (*S*s) to judge the emotional quality of stimuli presented too briefly and at illumination levels too low for correct specific identification. Two sets of words, previously judged to be pleasant or unpleasant, respectively, were presented tachistoscopically to a group of *S*s. After the exposure of each word, the *S*s were required to judge whether the word was pleasant or unpleasant as well as to guess what the word was. The judgments of pleasant or unpleasant that occurred in the absence of correct identification of the word were then examined for evidence of above chance accuracy in affective classification.

Since the *S*s might be expected to make the judgments of pleasant and unpleasant upon the basis of their affective responses, it would seem that both Osgood's theory and the concurrent response theory would expect the *S*s' affective judgments to show better than chance accuracy when the specific identifications of the stimuli were in error. However, if this affective discrimination occurs at lower levels of the nervous system and temporally prior to specific verbal identification as Osgood has suggested, then at any given illumination level between chance and 100% performance we should expect that the affective judgments would yield a more accurate classification of the stimuli than would the specific identifying responses. The concurrent response theory, on the other hand, would expect the thresholds or accuracy of the two response systems to be about equal.

## METHOD

Sixty English words, 30 pleasant and 30 unpleasant, were selected from among a list published by Young (1937). The words varied from three to nine letters in length and all were capable of ready definition by the *S*s. The 60 words used are listed in Table 1. During the first session of the experiment, the *S*s were individually given a list of the 60 words in random order and asked to rate each word on a seven-point scale in terms of whether the word tended to produce a pleasant or unpleasant feeling. They were instructed to use a rating of 1 for the most pleasant and a rating of 7 for the most unpleasant words. Emphasis was given to the feeling aspect of the word in an attempt to produce an affective rather than a cognitive judgment. One to two weeks after the first session, each *S* was recalled for another individual session during which his prerecognition affective judgments were obtained to the 15 words he had rated most pleasant and the 15 he had rated as most unpleasant. In selecting the words for an *S*, an attempt was made to match the pleasant and unpleas-

TABLE 1
LIST OF STIMULUS WORDS

| Pleasant Words | Unpleasant Words |
|---|---|
| 1. sweet | 1. spider |
| 2. happy | 2. bowels |
| 3. home | 3. pimple |
| 4. lovely | 4. belch |
| 5. love | 5. hate |
| 6. beautiful | 6. death |
| 7. music | 7. sorrow |
| 8. mother | 8. vomit |
| 9. fragrant | 9. retch |
| 10. dessert | 10. pain |
| 11. vacation | 11. stench |
| 12. soft | 12. rotten |
| 13. tranquil | 13. war |
| 14. summer | 14. harsh |
| 15. warm | 15. sickness |
| 16. flower | 16. belly |
| 17. sunshine | 17. horrible |
| 18. travel | 18. murder |
| 19. food | 19. snake |
| 20. beauty | 20. dirty |
| 21. melody | 21. slimy |
| 22. money | 22. gripe |
| 23. murmur | 23. cold |
| 24. spring | 24. bitter |
| 25. friend | 25. spit |
| 26. happiness | 26. blood |
| 27. smooth | 27. nasty |
| 28. dance | 28. sour |
| 29. sleep | 29. ugly |
| 30. jovial | 30. putrid |

ant words on the Thorndike-Lorge semantic count and on the number of letters.[2]

This second session was divided into three stages. During the first stage the *S*'s recognition threshold for five neutral practice words tachistoscopically presented was obtained. The second stage was a training or learning period. Here, five pleasant and five unpleasant words, randomly selected from the *S*'s two groups of 15, were presented 10 times each at subthreshold levels as determined by the five practice words. The *S*s were instructed that 10 words were to be shown, five pleasant and five unpleasant, but that the duration and illumination level of exposure would probably be such that they would seldom if ever be sure just what word was being exposed. They were further instructed that following each word exposure they were to attempt to judge whether the word shown was pleasant or unpleasant as well as to guess as to what they thought the word specifically was.

Each of the 10 words was then presented in random order at an exposure duration of .01 seconds and at an illumination level 10 volts below their mean recognition threshold for the five practice words. After each word exposure when the *S*'s judgment and guess had been recorded, he was told whether the word shown had been a pleasant or unpleasant one. When each of the 10 words had been presented at this initial subthreshold level, the exposure illumination was increased by 1 volt and the process repeated. This procedure was continued with the illumination level being increased each time by 1 volt until each of the words had been recognized or correctly guessed on two successive trials or until each of the words had been presented 10 times.

Following these training trials the *S*'s affective judgments and word guesses for the remaining 20 words (10 pleasant and 10 unpleasant) were obtained. The procedure followed was the same as during the training trials except that the *S* was never given information as to the correctness of his judgments. Again a constant exposure duration of .01 sec. was employed and the initial exposure illumination level was approximately 10 volts below his word recognition threshold obtained during the training trials. After each of the 20 words had been exposed in random order at the initial illumination level, the illumination was increased by 1 volt and the 20 words again exposed but in a different random order. This procedure was continued until 18 of the words had been correctly identified at two successive illumination levels.

*Apparatus.* The experiment was conducted in a room of controlled constant illumination. A Stoelting Model 21213 Projection tachistoscope was used, and the exposure illumination level was controlled by varying the voltage across the projector lamp by means of a variac. The voltage across the lamp varied in the range of 30 to 50 volts which gave an illumination range of .41 to 2.73 millilamberts as determined by a McBeth Illuminometer. Within this range a unit change in voltage is linearly related to log millilamberts. The stimulus words were typed in capital letters on opaque drawing paper with an electric typewriter using a carbon backing. Each word was then mounted between 35 mm. square pieces of glass for use in the tachistoscope.

The *S* was separated from the experimenter (*E*) by means of a 2′ x 4′ plywood partition. In the center of the partition at eye level was an opal glass screen upon which the stimulus words were flashed from the tachistoscope on *E*'s side of the partition. The projected word had letters approximately ¼″ in height.

*Subjects.* Seventeen *S*s, nine males and eight females, were obtained in response to a request for volunteers from undergraduate summer session classes at the University of Illinois. All were naive as to the real purpose of the experiment.

## RESULTS AND DISCUSSION

The results from the training session were analyzed first. For each *S* a four-fold chi square was computed for the relation between the affective classification of the 10 training words (five pleasant and five unpleasant) and the *S*'s affective judgments of these words on those trials prior to his specific identification of them. The chi square values obtained for each of the 17 *S*s are shown in Table 2. For 12 of the *S*s the relationship was positive; that is, the prerecognition judgments were correct more often than incorrect, and for one of these *S*s the value of chi square attained significance at or beyond the .05 level.

Since five of the *S*s obtained values of chi square that reflected a relationship in the opposite direction, the over-all significance of the results could not be determined by the usual method of adding the independent chi square values. Instead the Fisher-Pearson test

TABLE 2

CHI SQUARE VALUES FOR THE RELATIONSHIP BETWEEN AFFECTIVE WORD CLASSIFICATION AND PRERECOGNITION AFFECTIVE JUDGMENTS
(Training trials)

| Subject | $\chi^2$ | Subject | $\chi^2$ |
|---|---|---|---|
| 1 | .09 | 10 | .69 |
| 2 | 6.47* | 11 | (−) 2.51 |
| 3 | .61 | 12 | (−) 1.27 |
| 4 | .90 | 13 | .29 |
| 5 | 1.09 | 14 | .72 |
| 6 | .31 | 15 | (−) .04 |
| 7 | (−) 1.03 | 16 | .12 |
| 8 | .13 | 17 | .85 |
| 9 | (−) .51 | | |

Note.—The minus sign in parentheses indicates a negative relationship between affective classification and prerecognition affective judgments.

* Significant at or beyond .05 level.

[2] The frequency matching was actually unnecessary as indicated by an average within *S* correlation of .093 between log frequency of the word and number of exposures necessary for its recognition.

was used (Gordon, Loveland, & Cureton, 1952). This involves computing the exact probability for each obtained chi square value and then transforming them to $\log_e$. The sum of these transformed values is distributed as chi square with 2n degrees of freedom or, in the present case, 34 *df*. The application of this test to the data of Table 2 resulted in a summed value of 44.16 significant between the .10 and the .05 level.

While the results from the training session are suggestive that affective judgments tended to be correct in the absence of correct word identification, not too much confidence can be placed in this finding. Even if the probability figure obtained had been much lower, the fact that the *S*s were given knowledge of results after each judgment may have enabled them to learn to associate the responses of pleasant and unpleasant to some partial cue. For example, an *S* might remember that the last time he had had the impression or perception of a blurred word with a tall letter in the center the response "pleasant" was called correct. Thus when a similar incomplete perception occurred, he would again respond with the judgment "pleasant."

The data obtained during the judgmental phase of this experiment are not only free of this possible artifact but they also provide a more sensitive test of the accuracy of prerecognition judgments since a larger number of words was used. In this phase the *S*s' thresholds were determined for 20 additional words (10 pleasant and 10 unpleasant) with no feedback as to the correctness of their prerecognition affective judgments. Again for each *S* a four-fold chi square was computed between the affective classification of the words and the prerecognition judgments. The chi square values for the individual *S*s are shown in Table 3.

TABLE 3

CHI SQUARE VALUES FOR THE RELATIONSHIP BETWEEN AFFECTIVE WORD CLASSIFICATION AND PRERECOGNITION AFFECTIVE JUDGMENTS
(Judgmental trials)

| Subject | $\chi^2$ | Subject | $\chi^2$ |
|---|---|---|---|
| 1 | 9.55** | 10 | .85 |
| 2 | (−) .53 | 11 | 1.13 |
| 3 | .37 | 12 | .01 |
| 4 | 1.51 | 13 | .93 |
| 5 | .52 | 14 | 4.47* |
| 6 | (−) .22 | 15 | 14.63** |
| 7 | (−) 1.60 | 16 | (−) .51 |
| 8 | (−) .86 | 17 | 3.67 |
| 9 | 5.42* | | |

Note.—The minus sign in parentheses designates a negative relation between affective classification and prerecognition affective judgments.

* Significant at or beyond .05 level.

** Significant at or beyond .01 level.

For 12 of the 17 *S*s the direction of the relationship is positive, and for five of these *S*s the value of chi square is significant at about or beyond the .05 level. When the results for the group as a whole are evaluated via the Fisher-Pearson test, a chi square of 78.2059 is obtained, which with 34 *df* is significant beyond the .005 level.

In view of the evidence that *S*s could make prerecognition affective judgments of the stimulus words at better than chance accuracy, several other analyses were performed to insure that this result was not attributable to some obscure artifact. Each *S*'s prerecognition affective judgments were correlated with the number and kind of vowels contained in the different stimulus words as well as with consonants and number of letters. None of the obtained relationships were significant or appreciable, with the exception of the number of letters in the word. Here a small but consistent trend was found for *S*s to judge words with more letters as unpleasant. This effect, however, worked against our obtained result, since there was also a slight tendency for the longer words to be pleasant in classification.

The possibility that accuracy of prerecognition affective judgments was related to general perceptual acuity was also considered. The six *S*s with the highest and the six *S*s with the lowest prerecognition judgmental accuracy were selected. They were then compared on the voltage level where 50% and 90% of the words had been correctly identified. They were also compared on the number of times the stimulus words were exposed prior to reaching a level of 50% identification. The high accuracy *S*s identified 50% of the words at a mean voltage level of 45 compared with a mean of 43.5 for the low group. The resulting value of *t* (.91) could be expected more than 30 times in a hundred by chance. For the 90% identification level, the high group had a mean of 46.3 volts compared with a mean of 44.7 for the

low group ($t = .88$, $p > .30$). Negative results were also obtained for the number of exposures required to reach the 50% identification level, with the high group having a mean of 7.2 exposures compared with the low group mean of 6.8 ($t = .23$, $p > .30$).

The above analyses would certainly seem to bear out the conclusion that in the absence of specific recognition or identification of the stimulus words, the affective judgments tend to be right more often than wrong. This result is commensurate with Osgood's and Lazarus' theorizing as well as with the concurrent response model. However, these theories differ quite importantly in how they interpret or explain this result. Osgood (1957) and Lazarus (1956) assume that affective discrimination occurs temporally prior to specific identification during the process of perception. It follows from this assumption that affective discrimination has a lower threshold than does identification and that it is possible to select a level or magnitude of stimulation where the affective responses would show a significantly higher level of accuracy than the specific identifications.

The concurrent response explanation, on the other hand, assumes that identification and affective responses occur concurrently in the perceptual process and that the threshold for identification is at least as low as that for affective classification. The obtained result is attributed to noncorrelated errors between the identifying and the affective responses. At levels of stimulation where neither the specific identifications nor the affective responses are 100% correct, these noncorrelated errors yield a correct affective response on some of the trials when the identification is wrong. While this combination of responses may occur on any given trial, the situation may be reversed on the next trial, and over a series of trials at this stimulation level the two responses would be expected to show the same over-all accuracy.

A differential test of these two explanations can be made in the present experiment by comparing the accuracy of specific identification with affective judgments at a given voltage level of stimulus exposure. Accordingly, for each *S* the voltage level corresponding to an accuracy of approximately 50% in the specific identifications was determined. Table 4 shows

TABLE 4

COMPARISON OF ACCURACY OF SPECIFIC IDENTIFICATION AND AFFECTIVE JUDGMENTS AT THE ILLUMINATION LEVEL WHERE SPECIFIC IDENTIFICATION WAS APPROXIMATELY 50% CORRECT

| | Mean Correct Responses | Mean Correct Responses Adjusted for Chance |
|---|---|---|
| Specific identification | 10.8 | 10.3 |
| | | $t = .30$, $P > .50$ |
| Affective judgment | 15.2 | 10.4 |

the number of correct identifications the average *S* had at this level as well as the number of correct affective judgments.

In order to compare the differential accuracy of affective judgments and specific identifications at this level of exposure, it is necessary to make corrections for differences in chance success. For the affective judgments the chance level is 50%, but for the specific identifications the chance level is unknown. We have taken 5% as a conservative estimate of this unknown value.[3] In correcting for chance success we employed a formula previously used by Bricker (1955) for comparable data in which $Rc = R - W\ (Pr/Pw)$. Here $Rc$ is the corrected value, $R$ the number correct, $W$ the number wrong, $Pr$ the probability of a correct response by chance, and $Pw$ the probability of a wrong response by chance. The average of the corrected values is also shown in Table 4.

As can be seen from Table 4, there is no difference in accuracy between the specific identifications and the affective judgments when chance success is taken into account. Stating the problem somewhat differently, if an observer wished to determine whether the *S* had been presented a pleasant or an unpleasant word, the observer would be correct as often if he used the *S*'s specific identifications as he

[3] The 1 in 20 chance level seems quite conservative, since the *S* did not know what words were being presented. One to two weeks previously he had rated 60 words, about half of them pleasant and half unpleasant. Assuming that he could recall all of these words and relate them to the present session, his chance level for specific identification would be about 1 in 30 even assuming he used his affective classification of the stimulus to guide his guesses. However, even a chance level of 1 in 30 would be conservative, since examination of the *S*'s guesses revealed a high percentage of words that were not included in the original list.

would if he used the *S*'s affective responses. This finding is consistent with the concurrent response interpretation of our previous result but at variance with what woud be expected from the theories of Lazarus and Osgood.

## SUMMARY

The present study investigated the capacity of human *S*s to judge the emotional quality of stimuli presented too briefly and at illumination levels too low for correct specific identification. Two sets of words, previously judged to be pleasant or unpleasant, respectively, were presented tachistoscopically to a group of *S*s. After the exposure of each word, the *S*s were required to judge whether the word was pleasant or unpleasant as well as to guess what the word was. The judgments of pleasant or unpleasant that occurred in the absence of correct identification of the word were then examined for evidence of above chance accuracy in affective classification.

Analysis of the results revealed that in the absence of specific recognition or identification of the stimuli, the *S*s' affective judgments were more often correct than could be attributed to chance. No evidence was found, however, that the affective judgments had a lower discrimination threshold than specific identifications. These results were explained by assuming that the perceptual process concurrently activated both verbal identification responses and affective responses and that these two response systems had some noncorrelated sources of error.

## REFERENCES

BINDER, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62**, 119–129.

BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24–29.

BRICKER, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.

ERIKSEN, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, **69**, 625–634. (a)

ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80. (b)

ERIKSEN, C. W. Unconscious processes. In M. R. JONES (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. of Nebraska, 1958.

GORDON, M. H., LOVELAND, E. H., & CURETON, E. E. An extended table of chi-square for two degrees of freedom, for use in combining probabilities from independent samples. *Psychometrika*, 1952, **17**, 311–316

LAZARUS, R. S. Subception: Fact or artifact? A reply to Eriksen. *Psychol. Rev.*, 1956, **63**, 343–347.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. JONES (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. of Nebraska, 1957.

YOUNG, P. T. A study upon the recall of pleasant and unpleasant words. *Amer. J. Psychol.*, 1937, **49**, 581–596.

Received May 16, 1958.

# THE EFFECTS OF A MONETARY REWARD ON THE ACQUISITION OF AN IMITATIVE RESPONSE[1]

JOHN T. LANZETTA AND VERA T. KANAREFF

*Fels Group Dynamics Center, University of Delaware*

IN CONTRAST to early results reported by Miller and Dollard (1941), several recent investigations have found a low level of acquisition of an imitative response in adolescent and adult subjects (*Ss*) even when such responses are instrumental to success on the imposed task (Luchins, 1944; Schein, 1954; Bush & Mosteller, 1955). Building on suggestions by Schein (1954) and Luchins (1944), a recent study (Kanareff & Lanzetta, 1958) examined the hypothesis that such attenuation of an instrumental imitative response reflects *Ss'* anticipation of negative sanctioning for nonindependent behavior in certain "test-like" situations. The results, in general, were supportive: When expectations of positive social sanctions for imitation were induced, the level of acquisition of imitation positively increased with increasing instrumental value of the imitative response. When expectations of negative social sanctions were induced no such relationship between level of acquisition and instrumental value of imitation was obtained. Apparently the anticipation of social disapproval for imitation inhibited such responses even when they were instrumental to goal attainment (being "correct" on a judging task).

The present study represents an extension of this investigation. It employs the same three probabilities of reinforcement for imitation and two instruction-induced sets previously used, but under conditions of a monetary reward for being "correct." It seems plausible to interpret the depression of imitative responses under negative sanctions as representing a compromise between being correct and conforming to "the rules of the game." If so, such a compromise should be more difficult to effect if the reward for being correct were increased. Increasing the reward for a correct response (i.e., imitation) should increase the motivation for imitating when such responses are instrumental to goal attainment and thus should attenuate the inhibiting effects of sanctions against imitation.

In contrast to the no reward conditions, then, the introduction of a monetary reward should produce, under negative sanctioning conditions, a positive relationship between level of imitation and probability of reinforcement similar to that obtained, with no reward, for the positive sanctioning condition. In addition, a reward should enhance the positive relationship between level of imitation and probability of reinforcement when imitation is positively sanctioned.

## METHOD

The procedure employed is in all essential respects identical to that previously used (Kanareff & Lanzetta 1958). *Ss* in pairs, performed a series of pitch discriminations ostensibly for the purpose of improving existing methods for the identification of underwater craft. Communication between *Ss* was prevented by seating them in separated, partially enclosed booths. A small metal "response box" containing two response keys and a pair of signal lights was located in each booth. The *S* indicated his judgment on each pitch discrimination (whether the second tone was higher or lower than the first) by pressing the appropriate key on the response box. The two *Ss* were told they would give their judgments in a fixed order and that their "partner's" judgment would be indicated by the colored lights above his response keys. Since both *Ss* were instructed that their "partner" would give his judgment first, and since the response box lights in both of the booths were connected to a central control console, *E* could feed in a programmed sequence of "partner's" judgments to each *S*.

The correctness of *S's* choice was indicated by the presence or absence of a white light located on the front panel of his booth; this was controlled by *E* from the central console. In 60 pairs of tones (the critical trials), both the standard and comparison tones were the same. In 15 other pairs of tones the second tone differed in pitch from the standard. The pairs of different tones were randomly interspersed among the critical trials in order to reduce the possibility of *Ss* suspecting that in the majority of cases both tones in a pair were the same. A random sequence of the entire set of 75 pairs of tones was recorded on tape along with time signals and starting instructions. The tone pairs were presented to *Ss* over speakers located in the experimental room.

[1] We are indebted to Jo Anne Davis for her assistance in conducting the experiment and analyzing the data.

### *Experimental Conditions*

As in the previous study, the major independent variables were probability of reinforcement (instrumental value of imitation) and instruction-induced set. In the present study, however, a monetary reward was given for "success" on the task. This reward was in addition to the previously used reinforcement provided by a signal light when a judgment was correct.

*Ss* were provided with 75 poker chips at the beginning of the experimental session. The portion of the instructions describing the reward procedure was as follows:

> Although the discriminations [you are to make] are difficult, *most* of our subjects readily learn to make finer judgments with practice and get at least 80% of the judgments correct. Some people, however, can learn much more rapidly than others. Previous experience has also shown that it's tedious to make a large number of judgments at a single setting. To stimulate and maintain interest we've devised a payoff system. Since most people get 80% of the judgments correct we'll pay *ten cents* for each of the judgments that you get correct over and above the 80% made by most people. Use the poker chips in your booth to keep track of the number of judgments you bet correct. You can do this by moving one chip over to the side for every correct judgment. Each chip beyond the 80% mark will be redeemable for ten cents.

*Percentage of "partner's" judgments called correct (probability of reinforcement).* The sequence of judgments considered correct for the 60 pairs of identical tones and the corresponding three response sequences for the "partner" was randomized within blocks of 10 trials, with the restriction that there be an equal number of "higher" and "lower" judgments. The three "partner" response sequences form the three probabilities of reinforcement, .80, .50, and .20, i.e., 80% of the "partner's" judgments were correct in the first sequence, 50% in the second sequence, and 20% in the third sequence.[2] For the 15 pairs of different tones the partner always judged correctly, and the *S* was correct if he agreed with his "partner."

Thus, imitation or consistent agreement with the "partner" is instrumental to goal achievement in the .80 probability of reinforcement condition: it would produce 80% correct judgments on the 60 critical trials and 100% correct judgments on the 15 pairs of different tones. In terms of the payoff schedule, consistent imitation would result in 63 correct judgments, 15 of which would be rewarded at ten cents per judgment. Conversely, opposition, or consistent disagreement with the "partner" is the instrumental pattern in the .20 probability of reinforcement condition, and neither imitation nor opposition is effective in the .50 probability of reinforcement condition.

*Instruction-induced set.* The two instructions used were designed to induce different "sets" toward imitative behavior. Negative sanctions toward imitation were introduced in Instructions II by implying a test-like situation and by stressing "separateness" of judgments. The relevant portion of the instructions read as follows: "We are attempting to determine whether the judgments made by a group of people are more reliable, more accurate, than the judgments of one person alone. We are, in addition, interested in the problem of how many people are needed to obtain accurate judgments. The problem may be posed as follows: Suppose that two people *individually* judge the same thing. How much difference in accuracy is there in their *separate* judgments and are their *separate* judgments more accurate than their combined judgments?" In addition, the statement, "This will be a test of your ability to discriminate tones in that we will see how well you do in comparison with each other" was included near the end of Instructions II.

The relevant portion of Instructions I, which were designed to induce neutral or slightly positive sanctions toward imitation, were as follows: "We are attempting to determine how much of an actual difference is necessary under varying conditions of distraction in order for people to recognize these differences. A related problem is the development of effective means for improving judgments that are difficult to make. One possibility is that working in a group may improve one's ability to distinguish differences when conditions for making such judgments are poor."

### *Experimental Design*

The two sets of instructions and the three probabilities of reinforcement formed a 2 × 3 orthogonal design. Twenty-four female and 24 male volunteers from the undergraduate classes at the University of Delaware were randomly assigned in like-sexed pairs to the six cells of the design with the restriction that each cell contain an equal number of pairs of each sex.

The number of imitative responses (judgments indicating agreement with the partner) per block of 10 trials in the 60 critical trials is the major dependent variable. A questionnaire assessing the *S*'s aspiration level and estimates of his success, his partner's success, and of the amount of agreement with the partner was also used to provide data about *S*'s expectations in the experimental situation.[3]

## RESULTS

### *Imitative Responses*

Figure 1 presents separately for Instructions I and II the mean number of imitative responses per trial block for the three probabilities of reinforcement. Over the latter portion of the trial series there appears to be little difference between instructions: the mean proportion of imitative responses varies positively with probability of reinforcement for

[2] These probabilities of reinforcement are taken with respect to the 60 critical trials. Since the partner's judgments were always correct on the remaining 15 trials, the probabilities of reinforcement for the entire sequence are .84, .60, and .36.

[3] In general, analyses of the questionnaire data proved unenlightening, and they are omitted from the results section.

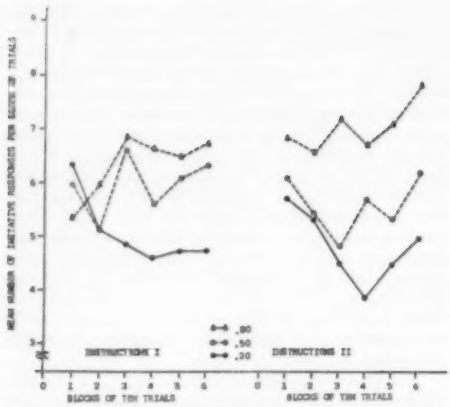


FIG. 1. MEAN NUMBER OF IMITATIVE RESPONSES PER BLOCK OF TRIALS AS A FUNCTION OF PROBABILITY OF REINFORCEMENT AND INSTRUCTIONS

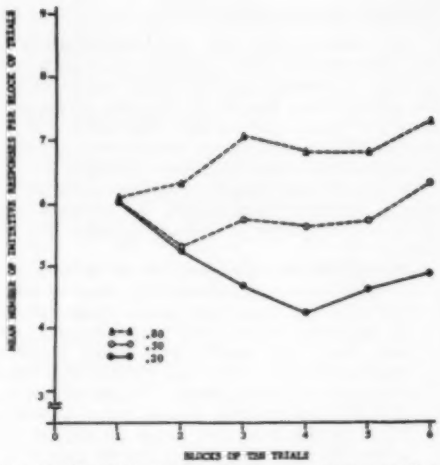


FIG. 2. MEAN NUMBER OF IMITATIVE RESPONSES PER BLOCK OF TRIALS AS A FUNCTION OF PROBABILITY OF REINFORCEMENT

both the negative and positive sanctioning conditions.

An analysis of variance of this data indicates, as predicted, no significant difference between instructions, nor is the instructions × probability of reinforcement interaction significant (Table A[4]). Probability of reinforcement and the interactions of probability of reinforcement with trials and with instructions and sex all contribute significant variance. Subjects within conditions are also significant indicating large individual differences in response to the experimental treatments.

The results of a Duncan Range test (Duncan, 1955) indicate that all three probability of reinforcement means differ significantly at the .05 level. The probability of reinforcement × trials interaction means are plotted in Fig. 2. The Duncan Range test results indicate that, in general, the three curves differ significantly from each other in the expected directions. However, for none of the reinforcement conditions does the level of imitation conform closely to the level of reinforcement, e.g., the .80 probability of reinforcement produces a maximum of 73% imitation. The Duncan Range test results for the instructions × probability of reinforcement × sex interaction means indicate that although there is considerable overlap in means, in general, the relationship between probability of reinforcement and level of imitation is more pronounced under negative sanctioning than positive sanctioning for males, while the converse is true for females (Table B[4]).

To return briefly to the relationship between level of imitation and probability of reinforcement, Estes (1957), on the basis of a number of experiments on probability learning, has advanced the notion that the cumulative proportions of a given response and the corresponding reinforcing event tend to equality as the number of trials approaches infinity. An examination of the present data plotted in terms of these parameters (Figs. 3, 4, and 5) indicates that such event matching, in general, does not occur.

As Fig. 3 shows, the cumulative proportion of the imitative response approaches but remains below the cumulative proportion of the corresponding reinforcing event $E_1$ for .84 probability of reinforcement condition.[5] It is

[4] Tables presenting data or analyses referred to in the text have been deposited with the American Documentation Institute. Order Document No. 5901, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] See Footnote 4.

[6] Probabilities of reinforcement for the entire sequence are .84, .60, and .36 as noted in Footnote 2. The .36 probability of reinforcement for imitation is equivalent to a .64 probability of reinforcement for opposition.

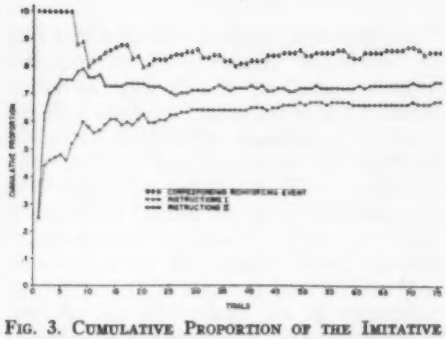

FIG. 3. CUMULATIVE PROPORTION OF THE IMITATIVE RESPONSE AND ITS REINFORCING EVENT IN THE .84 PROBABILITY OF REINFORCEMENT CONDITION

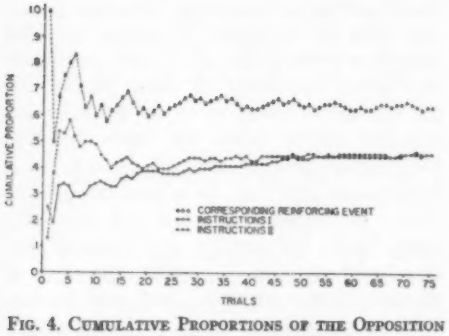


FIG. 4. CUMULATIVE PROPORTIONS OF THE OPPOSITION RESPONSE AND ITS REINFORCING EVENT IN THE .36 PROBABILITY OF REINFORCEMENT CONDITION

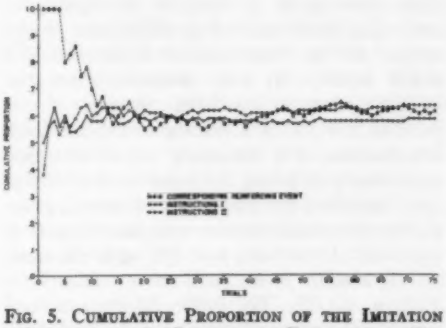

FIG. 5. CUMULATIVE PROPORTION OF THE IMITATION RESPONSE AND ITS REINFORCING EVENT IN THE .60 PROBABILITY OF REINFORCEMENT CONDITION

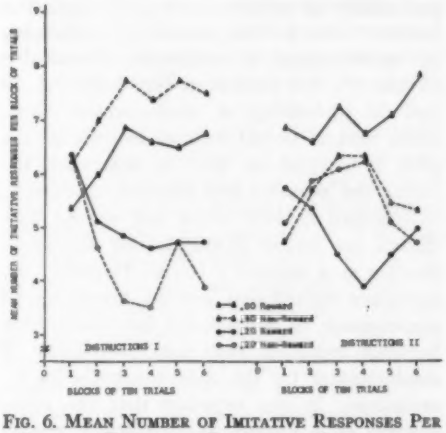


FIG. 6. MEAN NUMBER OF IMITATIVE RESPONSES PER BLOCK OF TRIALS AS A FUNCTION OF REWARD AND INSTRUCTIONS FOR THE .80 AND .20 PROBABILITY OF REINFORCEMENT CONDITIONS

also apparent that the correspondence between probability of reinforcement and level of imitation is a function of the instruction-induced set. From Fig. 4 it can be seen that a similar, but even greater, discrepancy between probability of reinforcement and level of opposition exists for the .64 probability of reinforcement condition, although here, the effects of instructions are much less pronounced. "Event matching" does occur, however, when the probability of reinforcement is .60, and under this condition instructions do not influence the level of imitative responses (Fig. 5). In terms of the present data, it appears that the extent of event matching depends upon the nature of the response and the conditions under which the response is elicited. Opposition responses are less modifiable by reinforcement than imitation responses and the frequency of both responses is affected by *S*'s set toward the experimental situation.

*Comparison with previous experiment.* Although an interval of approximately four months intervened between the present and previous study, there is little reason to expect any systematic bias to have affected the results of the two studies. Subject selection and assignment procedures were identical, and both studies were performed during approximately the same interval of the fall and spring semesters, respectively.

An analysis of variance of the combined data (Table C[7]) from the two studies shows

[7] See Footnote 4.

that once again $S$ variance is significant, indicating large individual differences in response to the experimental treatments. Of major interest in this comparison are the variance sources involving amount of reward as a variable. Contrary to expectations, introduction of a monetary payoff does not significantly increase the mean level of imitation. However, the interaction of reward, probability of reinforcement, and instructions is significant ($F = 8.36$, $p < .01$) as is the interaction, reward × instructions × trials ($F = 2.26$, $p < .05$). The latter interaction is of minor importance indicating only that both reward and instructions affect trial-by-trial variability in mean level of imitation over all probability of reinforcement conditions. The relevant data for the reward × probability of reinforcement × instruction interaction (Table D[8]) are plotted in Fig. 6 for the .80 and .20 probability of reinforcement conditions. Several points seem of interest in this plot. It is apparent that the differences between the negative and positive sanctioning instructions evident under low reward conditions are largely eliminated by the introduction of a monetary payoff. However, the monetary reward also has the unanticipated consequence of attenuating the relationship between imitation level and probability of reinforcement for the positive sanctioning instructions. It was expected that the introduction of a reward would produce greater congruence between probability of reinforcement and proportion of imitative responses for both the positive and negative sanctioning instructions.

One further observation on the combined results seems of importance. As previously noted, neither imitation nor opposition occurs at the level of the respective probabilities of reinforcement. However, the level of opposition is much more discrepant from the .64 probability level than the level of imitation from the .84 probability level. Opposition, in general, appears to be more difficult to condition than imitation, under all of the experimental conditions.

[8] See Footnote 4.

## DISCUSSION

The general hypothesis was advanced that the extent of utilization of an imitative response would vary directly with the instrumental value of imitation, inversely with the degree of negativeness of the prevailing social sanctions for imitation, and directly with increasing "utility" of the task goal. The rationale underlying this general hypothesis is based on a simple extension of the concepts of decision theory. Admittedly, given the difficulty of measuring most of the important parameters in the theory, the use of these concepts at the present time amounts to little more than a heuristic device.

In decision theory the important variables to be considered in predicting the results of a choice are the utilities of the outcomes associated with the available alternatives and the subjective probabilities of the outcomes ensuing. Given knowledge of these parameters, the organism chooses so as to maximize his expected utility. When the choice involves imitating or opposing another's decision in an achievement type task, we assume two general classes of outcome—a social and task outcome. To further simplify, it is assumed that the social outcome is either approval or disapproval of a reference group, and the task outcome is either success or failure with reference to the individual's level of aspiration. Further, the subjective probabilities associated with imitation or opposition for task outcomes are assumed to vary monotonically with the objective probability of reinforcement of the respective responses, while the subjective probability associated with imitation or opposition for social outcomes is assumed to be variable by instructions. In addition, the utility of task outcomes is expected to vary directly with reward, while the utility of social outcomes is a function of the expected rewarding or punitive measures administered by the reference group.

In the simplest situation of a single choice, where the possible alternatives are imitation or opposition, we assume the organism chooses that response with the highest expected utility. However, even gross prediction becomes more

equivocal when a series of choices must be made, and estimates of the probabilities of task outcomes must be deduced by *S* from the feedback from previous choices. In such cases we may talk about over-all strategies of imitation, opposition, or some mixture as constituting the class of possible response alternatives, in which case we would assume the organism chooses that *strategy* with the highest expected utility. However, in addition to the usual problems of measurement this imposes the additional requirement of estimating the utilities of all the possible combinations of imitation and opposition for a trial series—an impossible task at the present time.

The hypothesis advanced here represents, in a sense, an avoidance of the difficult measurement problems necessary for unequivocal prediction. Formulation of the problem in terms of the above variables, at the least, allows one to state what factors will affect the expected utility of a strategy. For example, we would expect the utility of consistent imitation to increase with increasing instrumental value of imitation (probability of reinforcement) and, contingent upon its instrumental value, to increase as the utility of task outcomes increases (amount of reward). We would also expect the utility of imitation to decrease as negative social sanctioning for imitation becomes intensified. This allows predictions of order for main effects and a crude sort of comparative prediction in those cases where incompatible influences are simultaneously varied.

The predictions of the present and past study are, generally, comparative in nature. The expected utility of imitation, and thus the extent of its use, should increase as the probability of reinforcement increases, but the increase should be greater for positive than negative sanctions. Thus we would predict a more pronounced positive relationship between level of imitation and probability of reinforcement under positive sanctions than under negative sanctions. In addition, the utility of imitation, when it is instrumental, should increase when a reward for "success" is introduced. A reward, therefore, should raise the level of imitation and overcome, in effect, the attenuation produced by the negative sanctions. The results of the studies, in general, confirm these predictions. When expectations of positive social sanctioning for imitation were induced, the rate and level of acquisition of imitative responses varied significantly with the probability of reinforcement of imitative responses. When expectations of negative social sanctions were induced, these differences in rate and level of acquisition were depressed (Kanareff & Lanzetta, 1958). However, as the present study indicates, the inhibiting effect of negative social sanctions is largely overcome by the introduction of a monetary reward. When the "utility" of "success" was increased, the rate and level of acquisition of imitation varied significantly with probability of reinforcement for both the positive and negative sanctioning conditions.

At a quite general level, then, the results support the theoretical rationale. However, there are several questions raised by the data which point up the necessity for quantitative measures of the intervening variables for any effective predictions from the theory. There is, for example, the large individual differences in response to the experimental treatments and the significant interaction of sex with instructions and probability ratio. Do these results indicate differences in the subjective probabilities of "success" associated with the alternatives, variability in the amount of anxiety associated with disapproval, or differing "utilities" attached to money as a reward. A plausible explanation for both individual variability and the sex interaction could be advanced in terms of any of these variables.

For example, it is possible that the implication of comparative evaluation contained in the negative sanctioning instructions, coupled with the tangible symbol of relative "success" (money) was differentially reacted to by the two sexes. Under such conditions the utility of "success" for males may be much greater than for females, and males would imitate relatively more than females when the "payoff" for imitation was increased.

On the other hand, the sexes may differentially evaluate success probabilities associated with imitation when a reward is introduced. When the stakes are high men may attempt to maximize their long-term payoff by figuring out a "system," whereas women may content themselves with being correct on some percentage of the single trials. Similarly, to account for individual variability, recourse can be had to differences in subjective probabilities, in utilities attached to social approval and/or disapproval, or in values associated with "success" and money as a reward. A choice between these, or other, alternative hypotheses can best be made when measures of such variables become available.

## SUMMARY

The present study extends the results of an earlier investigation of imitation and opposition to conditions of increased reward. The findings of the previous study supported the hypothesis that the extent of utilization of a response—i.e., imitation or opposition—was dependent not only on its instrumental value (probability of leading to reward) but also on the prevailing social sanctions for its use. In the present study it was hypothesized that the obtained attenuation of an instrumental imitation or opposition response under negative sanctions toward nonindependent behavior would be overcome by increasing the utility of achieving the task goal.

Two *Ss*, seated in separate booths, judged whether the second of a pair of identical tones, in a series of 60 paired tones, was higher than the first. Each *S* indicated his judgment by pressing the appropriate response key after his "partner's" choice was indicated by a light above the response keys; if *S*'s judgment was correct, a white light was flashed on, and, in addition, he moved a poker chip into a "win" box. The chips were redeemable for money at the end of the experiment according to a payoff procedure known to *Ss*.

Since each *S* was instructed that his partner would judge first, *E* was able to present a standard randomized sequence of "partner's" choices, as well as a prearranged sequence of reinforcements for imitation responses. The two independent variables were: (*a*) Probability of reinforcement for imitation. The percentage of the "partner's" choices called correct by *E* were either 80%, 50%, or 20%, i.e., the three probabilities of occurrence of reinforcement for imitative responses were .80, .50, and .20. (*b*) Instruction-induced set. One set of instructions was designed to induce expectations of neutral sanctions for imitation, while another set was designed to induce expectations of negative sanctions.

As predicted, the attenuation of instrumental imitative and opposition responses (i.e., imitation under .80 reinforcement and opposition under .20 reinforcement conditions) under negative sanctions obtained when the light alone was used as a reinforcement did not occur under the present monetary reward conditions. Apparently whatever inhibition was introduced by expectations of negative sanctioning for nonindependent behavior was overcome by increasing the payoff for successful performance. The introduction of a reward, however, had the unanticipated consequence of depressing the level of both imitation and opposition under neutral sanctioning conditions.

Under neither set of instructions or probability of reinforcement conditions (other than .50) did *Ss*' cumulative proportion of imitation or opposition responses match the cumulative proportion of the corresponding reinforcing event. Undermatching of the event occurred in the .80 and .20 conditions, the extent of the undermatching being determined by the social sanctions and amount of reward. In general, opposition was more difficult to condition than imitation under all experimental conditions.

Decision theory serves a heuristic purpose in discussing the results, and as a basis for suggesting variables that may account for the large individual variability and sex differences that were obtained.

## REFERENCES

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955. Ch. 12.

DUNCAN, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1–9.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

KANAREFF, VERA, & LANZETTA, J. T. The acquisition of imitative and opposition responses under two conditions of instruction-induced set. *J. exp. Psychol.*, 1958, **56**, 516–528.

LUCHINS, A. S. On agreement with another's judgment. *J. abnorm. soc. Psychol.*, 1944, **39**, 97–111.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

SCHEIN, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.

Received December 8, 1958. Prior publication.

# CRITIQUE AND NOTES

## PHANTOM EXPERIENCES IN MENTAL DEFECTIVE AMPUTEES

MARIANNE L. SIMMEL

*Duke University*

WHEN a normal adult person loses an arm or a leg in an accident or surgically, he typically continues to experience the extremity as if it were still present. This phenomenon, known for a long time, was given the name *phantom limb* by Mitchell (1871). Phantom experiences are not limited to limbs; they also occur after loss of such other body parts as nose, eyes, genitals, etc. Moreover, under some conditions, phantoms occur for extremities which, though physically present, have been deprived of sensation through blocking or transection of the neural pathways that ordinarily mediate the sensory input. The literature concerning this subject is extensive and detailed. But, as one attempts to survey both published and "common" knowledge, one quickly discovers a series of myths, published and unpublished statements, which apparently have never been put to a systematic test even though they may be rather crucial in interpreting the phantom experience. One such myth is that "mental defective amputees have no phantoms"—a statement which is frequently made outright in informal discussions and at least implied in occasional publications. The purpose of this paper is to present the findings of a small study of mental defective amputees.

### METHOD

Interviews were conducted with 26 mental defectives, each of whom had undergone some type of amputation at some time in his past. All but one of the patients were residents of one of five state schools;[1] in fact, they constituted the entire amputee population of these five institutions at the time of the investigation. The duration of the interviews ranged from a few minutes to as much as half an hour. Just as in studies of intellectually normal amputees, the patient was first asked to tell about the circumstances surrounding the loss of the arm or leg, when it happened, and how. While many normal amputees mention phantom experiences spontaneously when given an opportunity to talk about the amputation in some detail, this rarely occurred in the mental defectives. Almost invariably the writer had to lead off with a statement such as: "Some people, after they have lost a leg, say that they still feel it, even though they know it is gone. Has there been any time since the accident that you have felt as if the leg was still there, even though you knew very well that it was gone?" This highly suggestive question evoked positive responses from many patients, responses which could not be accepted at their face value. It was therefore followed up, either immediately or a little later, by a second question: "Some people, after they have lost a leg, never feel it any more; it is just gone, and they don't ever feel any part of it, not ever. How about you?" Some of the patients who had agreed to the first statement agreed just as lustily to the second statement. Some did not. Some started spontaneously to describe phantom toes or fingers, to point out that they felt only some parts and not others, to show on the contralateral healthy extremity the experienced position of the phantom or to relate incidents when they had "forgotten" the loss of the leg and "stepped" on the phantom.

From the hospital records, additional information was obtained regarding the age of the patient at the time of the interview, the time and cause of amputation plus any pertinent details, diagnostic category if available, educational achievement, and results of intelligence tests. Estimated mental ages were extrapolated for the time of the accident and the time of the interview.

### RESULTS AND DISCUSSION

Reasonably consistent evidence concerning presence or absence of phantoms could not be obtained at all in about one-half of the patients. This group includes, in the first place, a few patients whose general level of functioning was so low that any kind of communication was very difficult, let alone a conversation about such a patently unreal topic as phantoms. Second, it includes subjects with whom communication seemed more adequate, who appeared to stay with

[1] I am greatly indebted to the superintendents and the staff of these institutions for permission to examine the patients and for the many ways in which they cooperated in this study: Robert E. Wallace, Dixon State School, Dixon, Illinois; Joseph A. Albaum, Lincoln State School, Lincoln, Illinois; Richard C. Cooke, Myles Standish School, Taunton, Massachusetts; Karl V. Quinn, Wrentham State School, Wrentham, Massachusetts; and Clemens E. Benda, Walter E. Fernald State School, Waverly, Massachusetts. A special acknowledgement is due to Clemens E. Benda who arranged and facilitated most generously the examinations of the patients in the Massachusetts institutions.

the topic, but who agreed cheerfully to any suggestions made by the examiner and were totally unaware of the resulting inconsistencies. By contrast, evidence regarding the presence or absence of phantoms was judged to be reasonably reliable when it was given spontaneously during the conversation or in response to suggestive questions, provided it seemed to stand up under repeated though necessarily somewhat limited cross questioning.[2]

Furthermore, the possibility of confabulatory replies to suggestive questions must be considered, "wild stories" which might be maintained with some degree of consistency. It was hoped that, in the limiting case, such confabulations could be spotted, because the resulting descriptions would differ in significant aspects from the typical and lawful characteristics of the normal phantom.[3, 4] In fact, only two "atypical" descriptions were obtained—one of a phantom knee without foot or toes, and one of four phantom fingers without the thumb. Both of these came from patients whose reports were already classified as providing "no evidence" on the basis of apparent suggestibility and lack of consistency. None of the reasonably consistent reports contained unusual descriptions of phantoms.

Briefly, the 26 *Ss* fall into two groups: 14 patients from whom no evidence concerning presence or absence of phantoms could be obtained; 12 patients from whom reasonably consistent evidence was elicited. Eleven of the latter group reported phantoms, and one did not.

*Patients From Whom Reliable Evidence Could Not Be Obtained.* There were 14 patients from whom evidence concerning the presence or absence of phantoms could not be ascertained.[5] This group included three patients whose mental age at the time of the interview was below two years, i.e., patients with whom communication per se was impossible. Incidentally, their level of intellectual development at the time of amputation could hardly have resulted in phantoms, since amputation before chronological age of two years in normal children produces phantoms only very rarely (Simmel, 1958). The remaining 11 patients either did not seem to understand the questions at all, agreed enthusiastically to every question or statement made, or gave confused and conflicting reports.

TABLE 1

POSITIVE REPORTS OF PHANTOMS AS A FUNCTION OF EDUCATIONAL ATTAINMENT

| Highest Grade | No Evidence ($N$ = 14) No Phantom ($N$ = 1) | Probably Phantom ($N$ = 3) Phantom Definite ($N$ = 8) | Sum |
|---|---|---|---|
| (No info.) | (5) | (2) | (7) |
| <3 | 9 | 2 | 11 |
| 3+ | 1 | 7 | 8 |
| Sum | 10 (15) | 9 (11) | 19 (26) |

*Patients Giving Relatively Consistent Information.* From 12 patients reasonably consistent and reliable reports were obtained. Briefly, for eight of these 12 *Ss* there is very definite evidence of phantoms, though in one case it was transient, reported only by the patient's father and not remembered by the child herself. An additional three cases gave rather consistent evidence of at least intermittent phantom experiences, but since their descriptions were rather meager by comparison with the others, they were separately classified as "probably phantoms." Only one patient reported consistently that he had never experienced the absent hand in the 20 years since amputation.

*Relationships with MA, IQ, and Educational Achievement.* MAs and IQs at the time of amputation and at the time of the interview were much the same for the two groups, except for a small number of *Ss* at the idiot level who belonged necessarily to the first group. The educational attainment of the two groups was, however, quite different. Unfortunately, no information concerning schooling was available for seven of the *Ss*. As far as is known, only one of the 14 patients who could not give consistent evidence had reached the third grade, and at least five had never attended classes. By contrast, at least seven of the eleven patients

[2] For example: "How do you feel your fingers?"—"Like this, like a fist."—(later) "Can you move your fingers?"—"No."—"Are they spread apart like this?"—"No, they are like this" (makes fist with other hand).

[3] The typical characteristics of normal phantoms are well established in a large number of publications with detailed descriptions. For a recent summary cf. Simmel (1956a).

[4] This turns out to be a rather questionable assumption, since the very determinants of the distribution of the phantom might similarly determine the content of the confabulatory report. This problem is not at all analogous to that of the classical glove anesthesia versus loss of sensation due to neuronal lesion.

[5] Tables giving the findings in detail may be obtained from the American Documentation Institute, or they may be requested directly from the author. Order Document No. 5849 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

giving positive evidence of phantoms had reached the third grade and, as far as is known, only one never attended school. Table 1 shows the distributions. Omitting the subjects for whom there was no information, Fisher's test of exact probability yields a $p < .01$. While one can account for this relationship in several ways, it is suggested that a minimal level of cognitive stability and differentiation is a common prerequisite for (*a*) the ability to profit from primary grade instruction, especially with regard to reading and writing, (*b*) the development of phantoms following amputation, and (*c*) the ability to give reasonably consistent reports of an experience such as the phantom experience.

It may also be pointed out that the 11 patients with phantoms represent a "minimum incidence" in the group. Many patients asserted at one time or another that they did experience phantoms, but their reports were judged to be inconsistent and unreliable. Perhaps some of these patients did indeed have phantom experiences and simply could not relate them effectively. Furthermore, it so happened that the father of one patient was available and reported a very definite phantom soon after amputation, even though the patient herself did not remember or report such experiences. How much would the "incidence" have increased if an observant parent had been available for each patient in the group? It is not totally unlikely that phantoms develop even at quite low levels of cognitive maturity, but that they do not persist as long as in normal amputees, nor are they remembered by the defective patients.

SUMMARY

These findings dispose of the myth that mental defective amputees do not have phantoms. Of 26 *Ss*, there were 14 from whom no creditable evidence could be obtained, either because communication per se was impossible or because they agreed to everything said by the examiner. There were 12 who gave apparently reliable information; only one of these reported no phantom, the remaining 11 giving fairly consistent reports of phantom experiences. No relationship of phantom experiences with MA or IQ at the time of the amputation or the interview could be established. However, educational attainment was significantly greater for the *Ss* giving consistent and positive reports of phantoms than for those who did not.

REFERENCES

MITCHELL, S. W. Phantom limbs. *Lippincott's mag. pop. Lit. Sci.*, 1871, **81**, 563–569.

SIMMEL, MARIANNE L. On phantom limbs. *A.M.A. Arch. Neurol. Psychiat.*, 1956, **75**, 637–647. (a)

SIMMEL, MARIANNE L. Phantoms in patients with leprosy and elderly digital amputees. *Amer. J. Psychol.*, 1956, **69**, 529–545. (b)

SIMMEL, MARIANNE L. The conditions of occurrence of phantom limbs. *Proc. Amer. Phil. Soc.*, 1958, **102**, 492–500.

RIESE, W., & BRUCK, GABRIELLE. Les membres fantômes chez l'enfant. *Rev. Neurol.*, 1950, **83**, 221–222.

Received January 6, 1958.

# SOCIAL PERCEPTION AND THE MOTIVATIONAL STRUCTURE OF INTERPERSONAL RELATIONS[1]

MILTON E. ROSENBAUM

*State University of Iowa*

THE present investigation examines the effect on interpersonal perception of having individuals interact in pairs under instructions designed to require cooperation or competition. In a study that placed pairs of strangers in a "constructive interaction situation," Bieri (1953) found that participants viewed their partners as more similar to themselves after the interaction, while noninteracting pairs did not change in this respect. The interaction appeared to have no purpose for the discussants other than to satisfy the demands of the experimenter. It is conceivable that if the consequences of interaction were varied, "constructive" interaction might lead to increase in perceived similarity in some cases and to a decrease in others.

Previous studies employing a measure of assumed similarity have, for the most part, indicated that the degree of assumed similarity is a reflection of the attitude of the perceiver toward the object of his description. Fiedler, Blaisdell, and Warrington (1952) found that individuals perceive those

[1] This study was conducted while the author was at the University of North Carolina as part of the program of the Organization Research Group of the Institute for Research in Social Science. The Ford Foundation and the Office of Naval Research [Nonr-855 (04)] jointly support the O.R.G.

they like best as more similar to themselves than those they like least. Lundy, Katkovsky, Cromwell, and Shoemaker (1955) have shown that the aspects of oneself that are seen as desirable are attributed to liked persons, while unacceptable aspects are described as being possessed by disliked persons. Lundy (1956) finds that the assumption of similarity is made principally when the perceiver's attention is directed toward himself rather than to the other person.

Assuming that degree of perceived similarity reflects the affective attitude of the perceiver, several hypotheses are suggested. Deutsch (1949), who has dealt most elaborately with the consequences of cooperation and competition, hypothesized and found more positive feelings among group members under cooperative than under competitive instructions. Thus, within the present experiment, one would expect a greater change in the direction of increasing assumed similarity under cooperative than under competitive conditions. The experience of success rather than failure, moreover, should lead to more positive feelings toward others and, therefore, to a greater increase in perception of similarity. Riecken and Homans (1954) similarly hypothesize a positive correlation between "goodness" of activities and positive choice in interpersonal relations.

## METHOD

### *Subjects*

Fifty-two pairs of male undergraduates were recruited to appear in the small group laboratory to participate in an experiment which was to be counted toward their required experimental hours in the introductory psychology course. If *S*s, when they appeared, indicated that they knew each other beyond identification by name, they were not used in the experiment. Assignment to the particular conditions of the experiment was on a random basis.

### *Procedure*

About two weeks prior to the experiment, members of the course from which *S*s were recruited responded to a Self-Evaluation Form (SEF) which consisted of 30 trait pairs each separated by eight steps between the extremes. Examples of traits used are: "friendly–unfriendly"; "competent–incompetent"; "lazy–energetic"; etc.

At the time of the experiment, all pairs of *S*s received identical instructions, which indicated that the *S*s were to participate in a discussion that would be observed through a one-way screen mirror, the ostensible purpose being to study interaction patterns. Before receiving further instructions about the discussion, each *S* was asked to fill out a Social Sensitivity Schedule (SSS) as he believed his partner should have filled it out. The SSS was identical, except for heading and instructions, with the SEF. The *S*s were seated a distance from each other but after completion of the SSS were instructed to sit at a table facing each other.

Each pair of *S*s was then told to discuss the problem of choice of an ideal college or university for 20 minutes. They were told that their discussion would be evaluated by observers and that they might become eligible for a drawing of a pair of tickets to the conference basketball tournament. In half of the pairs, a cooperative set was induced by providing the information that each pair could only become eligible as a unit; one member could not be eligible without the other. A competitive set was induced in the other half of the pairs by informing them that if one member succeeded, then by exclusion the other failed.

The *S*s were then left to discuss the problem, while *E* joined the observers behind the screen. After 20 minutes, *E* interrupted and separated the *S*s by taking one *S* out to the anteroom. After apparently checking with the observers, *E* approached the appropriate *S*s individually and informed them that they had either succeeded or failed to become eligible for the tickets. Other *S*s were not informed of consequences. Assignment to a success, failure, or no consequence condition was on a random basis independent of any actual rating of interaction performance.

All *S*s were then asked to fill out the SSS once more as they now believed their partners should complete it. A questionnaire that included Back's Social Distance Scale (1951) was then administered.[2]

## RESULTS

*Effect on Assumed Similarity.* For each comparison between self description and the two descriptions of the partner, *D* scores as described by Cronbach and Gleser (1953) were computed, based on ratings of the 27 trait pairs that had positive and negative poles. If the rating of the partner was in the direction of a more favorable evaluation than that given to the self, a positive score was recorded. Change in *D* scores served as the measure of change in assumed similarity, with the direction of differences and of change taken into account.

[2] Back's scale asks *S* to indicate the distance he desires to maintain with his partner. He is asked to check *yes* or *no* to such statements as: "I would like to see him around campus some time"; "I would enjoy an animated discussion with him"; "I would discuss important personal problems with him." This is a seven-point scale developed by means of an abbreviated Thurstone technique.

TABLE 1

CHANGE IN DIRECTIONAL ASSUMED SIMILARITY AS A FUNCTION OF INSTRUCTIONS AND CONSEQUENCES

| Instructions | Adjusted Mean Change in D Scores | | |
|---|---|---|---|
| | Success | Failure | No Consequences |
| Cooperation | +2.19 | +2.27 | +3.54 |
| Competition | +0.26 | −0.16 | +1.99 |

Analysis of Covariance

| Source of Variation | *df* | Adjusted Mean Square | *F* | *p* |
|---|---|---|---|---|
| Instructions | 1 | 98.70 | 4.97 | <.05 |
| Consequences | 2 | 30.95 | — | |
| Interaction | 2 | 3.20 | — | |
| Within groups | 97 | 19.87 | | |

TABLE 2

SOCIAL DISTANCE AS A FUNCTION OF INSTRUCTIONS AND CONSEQUENCES

| Instructions | Mean Social Distance Score | | |
|---|---|---|---|
| | Success | Failure | No Consequences |
| Cooperation | 14.06 | 19.15 | 20.38 |
| Competition | 17.88 | 18.85 | 21.50 |

Analysis of Variance

| Source of Variation | *df* | Mean Square | *F* | *p* |
|---|---|---|---|---|
| Instructions | 1 | 62.50 | — | |
| Consequences | 2 | 200.25 | 3.67 | <.05 |
| Interaction | 2 | 32.30 | — | |
| Within groups | 84 | 54.50 | — | |

Because of a significant dependency of the directional change scores on initial rating reflected in an $r$ of .28, an analysis of covariance controlling initial rating was performed. The summary of this analysis and the adjusted means for change in directional assumed similarity appear in Table 1.[3] The difference between cooperative and competitive conditions (Instructions) is significant at beyond the .05 level.[4] *Ss* showed more change under cooperative than under competitive conditions in the direction of assigning more positive ratings to their partners than they did to themselves.

Although the over-all effect of the consequence variable did not approach a level of acceptable significance, the comparison of the effects of presentation of a consequence, be it success or failure, with the condition in which *S* was left without awareness of consequences, results in a $t$ ratio of 1.92 ($p < .06$).[5] The $t$ ratio for the difference between the success and failure conditions is .34 ($p > .70$). Evidence is therefore lacking in support of the hypothesis that success leads to a different evaluation than failure, but there appears to be a trend for *Ss* to perceive more positive qualities in their partners following interaction in which no feedback as to consequences is offered than when such information is provided.

*Effect on Social Distance.* Previous studies have indicated a positive relationship between assumed similarity and sociometric choice (Fiedler et al., 1952). Back's scale was administered in order to assess this relationship within the present setting. It was scored by assigning values 1 to 7 to the items and adding the scores for items checked *yes*. The higher the score, the closer the relationship desired. The correlation obtained between Back's scale and the scores reflecting change in perceived similarity was +.04.

The scores from Back's scale were also evaluated in terms of the experimental conditions (Table 2).[6] In an analysis of variance, summarized in Table 3, the cooperation–competition difference was not significant, while the consequence variable was significant at beyond the .05 level. A comparison of the two conditions in which the *Ss* were made aware of consequences with the remaining condition in which no consequence was reported indicated that this difference ($p < .05$) accounts for the significant main effect. There was no difference between success and failure conditions. These findings resemble the trend found in the analysis of the data concerning assumed similarity.

## DISCUSSION

On the supposition that cooperation constitutes constructive interaction while competition does not, the findings tend to support Bieri's contention

[3] Bartlett's test of homogeneity of variance was employed for all analyses and indicated support for the assumption of homogeneity.

[4] Analysis of data involving a measure of assumed similarity that did not take directionality into account yielded similar results but at slightly less than an acceptable significance level ($p = < .07$).

[5] With 102 degrees of freedom, a $t$ of 1.96 is required at the .05 level of significance.

[6] Because of failure on the part of several *Ss* to complete the questionnaire, a reduced $N$ was considered. The score of one *S* was randomly removed to achieve proportionality among the cells.

that constructive interaction between partners leads to an increase in assumed similarity. Despite earlier expectations, no support is given to the hypothesis that the specific consequences of the interaction affect the perception of similarity. If assumed similarity reflects affective orientation, this finding is surprising in the light of several studies (Filer, 1952; Gebhard, 1948) that report consistent results to the effect that success leads to an increase in positive orientation, while failure has the opposite effect. Although previous studies have for the most part dealt with the evaluation of goal objects, Gilchrist (1952) found that the experience of success or failure influenced interpersonal orientations. In the present experiment the *S* could not know whether to attribute success or failure to his own efforts, to his partner's actions, to both in some weighted combination, or to some undecipherable whim of *E* and his "observers." If atrribution of causality can be considered a relevant basis for interpersonal evaluation (Heider, 1944), the large variance contributed by the suggested sources might easily nullify differences ordinarily expected to be created by attribution to any one source.

The failure to find a significant relationship between the measures of assumed similarity and social distance is similar to a finding of Lundy (1954). He attributed lack of correlation to the shortness of the contact. A similar hypothesis may be invoked here.

A provocative finding concerns the differences between the conditions in which knowledge of consequences was provided and the condition in which it was withheld. The data suggest that greater positive change in evaluation follows interaction when no information is provided as to success or failure. Further research is necessary to ascertain the stability of this finding and to clarify its origins.

## SUMMARY

Pairs of college males, unacquainted with one another prior to the experiment, participated in a 20-minute discussion under conditions of either cooperation or competition with anticipation of an attractive prize for a good discussion. Following the discussion, the *S*s were told, on a random basis, that they succeeded or failed, while in a control condition they were not informed of consequences. On three occasions, the *S*s filled out forms descriptive of personal characteristics. On the first occasion, *S*s described themselves. Prior to and following the discussion, *S*s described their partners. Difference scores based on changes in perception of similarity of partner to self served as a principal measure. Back's Social Distance Scale was also administered.

Analysis of covariance of difference scores indicated in the primary analysis only one significant effect. *S*s under cooperative instructions consistently saw their partners as more similar to themselves after interaction than before than did *S*s under competitive instructions. In terms of Back's scale, *S*s desire to maintain a greater social distance between themselves and partners when knowledge of the consequences of the interaction is provided than when no such information is available.

## REFERENCES

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **64**, 9–23.

BIERI, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, **48**, 61–66.

CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.

DEUTSCH, M. An experimental study of the effects of cooperation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199–232.

FIEDLER, F. E., BLAISDELL, J., & WARRINGTON, W. G. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.

FILER, R. J. Frustration, satisfaction, and other factors affecting the attractiveness of goal objects. *J. abnorm. soc. Psychol.*, 1952, **47**, 203–212.

GEBHARD, MILDRED E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation, and need. *J. exp. Psychol.*, 1948, **38**, 371–378.

GILCHRIST, J. C. The formation of social groups under conditions of success and failure. *J. abnorm. soc. Psychol.*, 1952, **47**, 174–187.

HEIDER, F. Social perception and phenomenal causality. *Psychol. Rev.*, 1944, **51**, 358–374.

LUNDY, R. M. Assimilative projection in interpersonal perception. Unpublished doctoral dissertation, Ohio State Univer., 1954.

LUNDY, R. M. Assimilative projection and accuracy of prediction in interpersonal perceptions. *J. abnorm. soc. Psychol.*, 1956, **52**, 33–38.

LUNDY, R. M., KATKOVSKY, W., CROMWELL, R. L., & SHOEMAKER, D. J. Self acceptability and descriptions of sociometric choice. *J. abnorm. soc. Psychol.*, 1955, **51**, 260–262.

RIECKEN, H. W., & HOMANS, G. C. Psychological aspects of social structure. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 786–832.

Received January 23, 1958.

# REGRESSION UNDER STRESS TO FIRST LEARNED BEHAVIOR[1]

RICHARD P. BARTHOL[2]
*University of California, Los Angeles*
AND NANI D. KU
*Pennsylvania State University*

IN A REPORT on studies of frustration in young children, Barker, Dembo, and Lewin (1941) and Barker, Dembo, Lewin, and Wright (1947) state that "... strong frustration causes tension which leads to emotionality and restlessness, to de-differentiation of the person, and hence to behavioral regression." They described behavior changes in children when faced with a frustrating situation, indicating less "constructiveness of play," a deterioration of social interaction, and "intellectual regression." Other studies (Hamilton & Krechevsky, 1933; Kleemeier, 1942; O'Kelly: 1940a, 1940b) have accumulated evidence showing that there is regression in behavior following stress or frustration.

This paper describes a study designed to test the hypothesis that under stress or frustration the person regresses to the *earliest* learned behavior that is appropriate to the situation. In other words, if a person has learned two or more responses to a stimulus situation, he will respond with the first learned of these when placed under stress. Mowrer (1940), among others, has shown that frustrating a particular learned behavior causes the organism to revert to earlier learned behavior. Maier (1949, p. 66) has pointed out some of the problems inherent in these earlier studies:

> Hamilton and Krechevsky, Everall, Sanders, Martin, O'Kelly, and Mowrer believe that frustration causes a return to a former response (historical regression), but in their studies adequate controls were not present for distinguishing between the return to a former response and the abandoning of the prevailing response for a new one that just happened to correspond to an earlier response. The above authors used punishment for disrupting a prevailing response and demonstrated that some animals persisted in this response whereas others did not. That frustration tends to make animals return to a former response, however, is open to question.

In the view advocated here, the source of stress may be entirely unrelated to the behavior, and the person may have a free choice among behaviors that were learned at various times. The terms "stress" and "frustration" are used interchangeably under the assumption that frustration, fatigue, and conflict are specific cases of the more generalized state of stress.

## METHOD

The tying of a bowline knot was selected for study because there are two quite dissimilar methods of achieving the end result. Eighteen *S*s were randomly divided into two groups. Both groups were taught both methods of tying the knots but in different order. In a later stressful situation *S*s were asked to "tie a knot." Whether the *S* used the method learned first or second was then noted.

The details are as follows. The *S*s, 4 men and 14 women undergraduates at the Pennsylvania State University, were told that they were participating in two separate studies. One was an investigation of training methods, and the other was the validation of a new intelligence test. Ostensibly as part of the former study they met with the second author to learn how to tie a bowline. Each *S* learned both methods of tying the knot. There is a simple straightforward way of tying a bowline, Method A, that includes a story which makes it easy to learn. Method B is more elegant and faster but harder to learn. Most *S*s seemed to prefer tying the knot by Method B once they had mastered it. The motions used for the two methods are quite dissimilar; until the final product is achieved there is no apparent relation between the two methods.

Group I learned Method A first; Group II, Method B. The training sessions were conducted informally and there was no evidence of stress or anxiety. The criterion for training was five consecutive perfect trials, each trial under four seconds. After a lapse of time ranging from one to four days, each *S* was trained in the second method of tying the knot. The time between training periods was determined by the mutual convenience of *E* and the *S*s. All *S*s completed training no later than three days prior to the stress situation.

The stress situation consisted of the taking of a difficult intelligence test late at night under restrictive conditions. The first eight *S*s, all women, reported to *E* at 1:00 A.M. on a Sunday morning immediately after returning from a dance. The remaining ten, four men and six women, reported

[1] This paper was reported in part at the 1955 meetings of the American Psychological Association.

[2] The experiment was conducted while the writer was on the faculty of the Pennsylvania State University.

to *E* immediately after finishing a final examination at 9:00 P.M. on a Thursday evening.[3] The *S*s were told that they were to take a simple intelligence test designed for high school students, and that the average college student would be able to answer most of the questions without difficulty. Actually the test used was the Lepley Synonym Vocabulary Test (Lepley, 1955; Lepley & Zeigler, 1956), which is designed to discriminate at the upper levels of intelligence as well as at the normal range. The average college student would find many words in the test that he would not have encountered before. The test has 16 parts, each of which requires the *S* to match each of 30 words with one of five common words. The test usually requires between 40 minutes and one hour for completion. The *S*s were required to spend 12 minutes on each part and were not allowed to proceed to the next section until the time had elapsed. The total time was thus 3½ hours. The *S*s had time after each section during which they had nothing to do. They were not allowed to smoke, talk, leave the room, or direct any overt activity toward anything but the test proper. Notes made by *E*s indicate that the experience was stressful. Seven of the *S*s had to be kept awake by constant reminders. One *S* burst into tears. In spite of the admonitions for silence many comments were made such as "how many more minutes to go" or "can't I smoke just one cigarette?" At the end of the time the group was extremely irritable and restless. It should be pointed out that the *S*s knew that they would be kept up late and that they were to take the test. After the conclusion of the experiment a complete explanation was given, and the *S*s departed quite amicably.

At the end of the 3½-hour test, the *S*s were called out of the room individually in random order, handed a cord, and told "tie a knot."

As a check on the importance of the stress situation in determining the response, all *S*s had been asked to tie a knot under conditions that were considered to be without stress. The results were that nine tied the knot by the method learned first and nine by the method learned second. Method B was slightly preferred, ten to eight. These results indicate that the factor of primacy was not the determining one in a nonstress situation.

## RESULTS

Sixteen of the 18 *S*s tied a bowline by the method learned first. The remaining two tied a bowline by the method learned second. Using the binomial test (Siegel, 1956, p. 39), the null hypothesis was rejected at the .001 level.

All of the nine *S*s in Group I used Method A; seven of the nine in Group II used Method B. One of the two who used Method A volunteered the fact that he had belatedly remembered that he had learned to tie the knot in this way while a Boy Scout. (All volunteers had been screened for this information before the training period; two had been rejected because they had already learned the knot.) The other *S* in Group II who used Method A started to tie the knot by Method B but stopped midway and used the other, simpler, method.

## DISCUSSION AND CONCLUSION

The initial hypothesis was clearly supported for the case of *two* learned responses: if a person has learned two alternate responses to a stimulus and is placed under stress unrelated to the behavior being observed, he responds to the stimulus with the earlier learned behavior pattern. The more general formulation that regression occurs to the *first* learned behavior appropriate to the situation was not adequately tested, since there were only two responses available to the *S*s; a study allowing three or more alternatives would serve to decide whether the response pattern is the first learned or merely an earlier learned response.

The stimulus, "tie a knot," was deliberately chosen to allow the *S*s the opportunity of tying a knot that had been learned under different conditions and much earlier than the bowline. The second author, who had trained all of the *S*s in the knot tying, and who therefore was presumably identified with the bowline, gave the *S*s the stimulus instructions. All of the *S*s were familiar with the simple overhand knot, usually learned at a very early age while learning to tie shoes. Not only was this response learned earlier, but it had received much more practice over the years and had the advantage of recency, since the *S*s had all tied their shoes subsequent to the training on the bowline. No *S* tied an overhand knot or any other knot except the bowline. This result was predicted on the grounds that tying the bowline would be seen as the only appropriate behavior under the circumstances.

The concept of specific regression suggests that under stress the person does not simply go back along some time dimension and produce more "primitive" behavior as if he were younger, but instead responds in a way that gave him his first taste of success under similar conditions. Lepley (1954) has stated that variability in behavior is decreased under disturbed and excited states, a statement that fits with the concept of regression under stress. He also indicated that variability has

[3] Separate analyses of the results for the first eight and the last ten *S*s support the same conclusions as the pooled analysis. Only the latter will therefore be reported in this paper.

high survival value and is related to intelligence. If regression is not adaptive because it tends to reduce variability of behavior, one might ask why the evolutionary processes have not weeded out this characteristic. Perhaps it is because at times of stress the organism does not have time to consider what it is going to do, but instead must act immediately. Other factors then determine whether the response is adequate to meet the needs of the situation. Thus, even stereotyped or regressive behavior allowed Aesop's cat to escape the hunters, since it was not necessary for him to take the time to select one of several responses. The fox, with many available responses, was caught and killed while making up his anthropomorphic mind.

## REFERENCES


BARKER, R. G., DEMBO, T., & LEWIN, K. Frustration and regression: An experiment with young children. *Univ. Ia. Stud. Child Welf.*, 1941, **18**, No. 1.

BARKER, R. G., DEMBO, T., LEWIN, K., & WRIGHT, M. E. Experimental studies of frustration in young children. In T. M. Newcomb & E. L. Hartley (Ed.), *Readings in social psychology*. New York: Holt, 1947. Pp. 283–290.

HAMILTON, J. A., & KRECHEVSKY, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, **16**, 237–253.

KLEEMEIER, R. W. Fixation and regression in the rat. *Psychol. Monogr.*, 1942, **54**, No. 4 (Whole No. 246).

LEPLEY, W. M. Variability as a variable. *J. Psychol.*, 1954, **37**, 19–25.

LEPLEY, W. M. The rationale, construction, and preliminary try-out of the Synonym Vocabulary Test. *J. Psychol.*, 1955, **39**, 215–225.

LEPLEY, W. M., & ZEIGLER, M. L. The Synonym Vocabulary Test: Standardization and validation. *J. Psychol.*, 1956, **41**, 419–425.

MAIER, N. R. F. *Frustration: The study of behavior without a goal*. New York: McGraw-Hill, 1949.

MOWRER, O. H. An experimental analogue of "regression" with incidental observations on "reaction formation." *J. abnorm. soc. Psychol.*, 1940, **35**, 56–87.

O'KELLY, L. I. An experimental study of regression. I. Behavioral characteristics of the regressive response. *J. comp. Psychol.*, 1940, **30**, 41–53. (a)

O'KELLY, L. I. An experimental study of regression. II. Some motivational determinants of regression and perseveration. *J. comp. Psychol.*, 1940, **30**, 56–95. (b)

SEARS, R. R. Survey of objective studies of psychoanalytic concepts. *Social Sci. Res. Council*, 1943, **51**, 76–104.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.


Received January 7, 1958.

# THE AFTERIMAGES OF "HALLUCINATED" AND "IMAGINED" COLORS[1]

THEODORE XENOPHON BARBER

*Harvard University*

DOES the hypnotic subject "actually experience" the suggested "hallucinatory" object or does he just state that he sees it in order to please the hypnotist? Approaching this question indirectly, by suggesting the hallucination of colored forms to their hypnotic subjects, Erickson and Erickson (1938) and Rosenthal and Mele (1952) found that "hypnotically hallucinated" colors have at least one property in common with "actually experienced" colors: they are consistently followed by complementary colored afterimages. However, these investigators did not attempt to determine if subjects can "see" the appropriate afterimages or "hallucinated" or "imagined" colors *without* a "hypnotic induction procedure." In fact, they seem to agree that, to carry out this task, an individual must be given not only a lengthy "hypnotic induction procedure" but also a period of "at least 15 to 30 minutes [after a 'trance' has been induced] to permit the development of possible neuro-and-psycho-physiological changes and to adjust himself" (Erickson, 1941). The following experiments were therefore designed to answer two questions:

1. Will some hypnotic subjects, who are unable to define or name complementary colors, report the appropriate afterimages of "hallucinated" colors immediately after a "minimal" (i.e., less than five minutes) "hypnotic induction procedure?"

2. Will some individuals, who are unable to define or name complementary colors, report the appropriate afterimages of "imagined" colors *without* a "hypnotic induction procedure?"

## EXPERIMENT 1

### *Procedure*

Six male college students who had previously "experienced" a number of "hypnotic phenomena" —e.g., age-regression, negative hallucinations,

[1] This research was supported by postdoctoral fellowship, MF6343c, from the National Institute of Mental Health of the Public Health Service.

amnesia, complex posthypnotic behavior—were given a modified hand levitation hypnotic induction procedure as described by Wolberg (1948) but limited to a period of less than five minutes. Immediately thereafter the subjects (*Ss*) were instructed:

> In a moment I shall show you a blue circle. Keep looking at the blue circle until I remove it and show you another circle. Then keep looking at the second circle and tell me what you see there.

A circle of one-inch radius had previously been inscribed (by pencil) on a 5″ by 8″ *white* card. This card was presented approximately 24 inches from the *S*'s eyes and held there for 30 seconds. The experimenter (*E*) repeated, "Keep looking at the blue circle." The card was then removed, and a similar white card with a similar circle was presented for 15 seconds. *E* repeated, "Look at this circle and tell me what you see there." After the *S*'s report, he was given another white card with an inscribed circle and told to look at the "red" circle. This card was also removed after 30 seconds, and another white card was substituted for 15 seconds. The procedure was the same for "yellow," "green," and "orange."

The *Ss* who reported afterimages in the complementary colors were then asked:

> What did you expect to see when you looked at the second card after you had been looking at blue (and then red, yellow, green, and orange) for about half a minute? . . . You have heard of afterimages and complementary colors, haven't you? Tell me what you know about them. . . . Why do you think you saw the color yellow after you looked at blue (and then green after looking at red, etc.)?

The *Ss* who did not report the appropriate afterimages of the suggested colors were asked, "Tell me what you really saw when I told you to look at the blue (red, yellow, green, and orange) circles."

After answering these questions the *Ss* were "given amnesia" for "the hypnosis" and "awakened."

*Results*

Four of the six *Ss* did not "hallucinate" the suggested colors; they reported that they saw "only white circles." However, two of the *Ss* stated that they "saw" the (suggested) colors and, as Table 1 indicates, they consistently reported the appropriate negative afterimages.

When questioned about their knowledge of afterimages, the two *Ss* who had reported afterimages in the complementary colors denied prior knowledge of afterimages, were unable to define or name complementary colors, insisted that they had expected to see the *same* color when they looked at the second card, and stated that they "saw" the first (suggested) color and the second (complementary) color simply because "the colors were just there."

TABLE 1
AFTERIMAGES OF "HALLUCINATED" COLORS DURING "HYPNOSIS"

| Subject | "Hallucinated" Color | Afterimage Report by *S* |
|---|---|---|
| JC | Blue | Yellow |
| JH | | Yellow |
| JC | Red | Blue |
| JH | | Green |
| JC | Yellow | Blue |
| JH | | Violet |
| JC | Green | Purple |
| JH | | Red |
| JC | Orange | Blue |
| JH | | Sky blue |

The four *Ss* who did not "hallucinate" the colors stated that they "knew," during the experiment, that the cards were white; e.g., "I figured that you wanted me to think that the cards had colors but I realized they were all white."

## EXPERIMENT 2

*Procedure*

Approximately one-half hour after the preceding experiment, the same *Ss* participated in a second experiment in which they were given the following instructions:

> I am not going to hypnotize you in this experiment. Instead, I am going to ask you to do something which you may find difficult. Keep looking at the white circle on this card until you are able to make it appear blue. I'm certain that you can do this. Try very hard to imagine the color blue and then project the imagined color on the circle. Let me know when you have made the white circle appear blue.

One of the white cards that had been used in the preceding experiment was presented approximately 24 inches from the *S*'s eyes. When the *S* stated that he had projected a blue color on the inscribed circle, he was asked to continue looking at it; after 30 seconds the card was removed, and a similar card with a similar circle was presented for 15 seconds. *E* stated, "Keep looking at the

circle on this card and tell me what you see there." The same procedure was used for "red," "yellow," "green," and "orange."

After completing the experiment, the *S*s who had reported the appropriate afterimages of the "imagined" colors were asked essentially the same questions as in the preceding experiment. After answering the questions, the *S*s were tested for afterimage responses to each of the *real* colors.

*Results*

Four of the six *S*s reported (within two to four minutes) that they "just couldn't do it" or that "it was impossible to project a color in this way." However, two *S*s reported that they had "projected" the imagined color on the white circle and, as Table 2 indicates, they consistently reported the appropriate afterimages of the "projected" colors. (With each of the five colors, these two *S*s required between 2 to 5 minutes before they stated that they had succeeded in "projecting" the color on the white card.) The two *S*s who reported the appropriate afterimages of the "imagined" colors during this "waking" experiment were *not* the same *S*s who reported complementary colored afterimages of the "hallucinated" colors during the previous "hypnotic" experiment.

When questioned as to their knowledge of afterimages, one of the two *S*s who had reported the appropriate afterimages was unable to define or name complementary colors and insisted that he did not expect to see "anything" when he looked at the second card. The second *S* was able to define and name complementary colors; however, he insisted that his prior knowledge did not affect his responses and that he was "surprised" when he saw the negative afterimages.

These two *S*s also reported that they felt as if they had "gone into trance (or hypnosis)" when they were "imagining" and "projecting" the colors. When asked, "What do you mean by 'going into trance (or hypnosis)?' " they answered, "I had to really concentrate on the color. I wasn't aware of much else," and, "I felt almost like I do in trance. I wasn't thinking about anything except getting the darn color on the paper."

When shown the real colors, all of the *S*s reported the appropriate negative afterimages.

TABLE 2

AFTERIMAGES OF "IMAGINED" COLORS DURING "WAKING"

(GB and JB participated in Experiment 2; PM and SS participated in Experiment 3)

| Subject | Colors *S* Was Asked to "Imagine" | Afterimage Reported by *S* |
|---|---|---|
| GB | Blue | Yellow |
| JB | | Yellow |
| PM | | Butter yellow |
| SS | | Gold |
| GB | Red | Greenish violet |
| JB | | Traffic light green |
| PM | | Bluish |
| SS | | Green |
| GB | Yellow | Bluish green |
| JB | | Blue grey |
| PM | | Greenish-blue |
| SS | | Greeny-blue |
| GB | Green | Pinkish |
| JB | | Pink |
| PM | | Nut brown |
| SS | | Red |
| GB | Orange | Blue |
| JB | | Greenish blue |
| PM | | Pale blue |
| SS | | Purple |

## EXPERIMENT 3

The *S*s of the preceding experiments were "good" hypnotic *S*s and had previously participated in a number of hypnotic experiments with the *E*. A third experiment essentially repeating the procedure of the second was performed to determine if *S*s with whom *E* did *not* have a subject-hypnotist relationship would report the appropriate afterimages of "imagined" colors. Six male and five female college students participated as *S*s under normal waking conditions.

Nine of the 11 *S*s reported (within two to three minutes) that they were not able to "project" the color on the white circle. The other two *S*s, however, stated (within three minutes) that they had "projected" the "imagined" color, and as Table 2 indicates, they consistently reported afterimages in the complementary colors.

Both *S*s who reported complementary colored afterimages denied prior knowledge of afterimages and were unable to define or name complementary colors. One *S* insisted that "there must be some trick in it" and the other *S* stated that "it must have something to do with trying to make a white circle look colored."

When asked to comment on their experience, these two *S*s stated, "It took a heck of a lot of concentration to get the colors on the circles," and, "I had to put a great deal of effort into it. I had to give it all of my attention."

When shown the real colors, all of the *S*s reported the appropriate afterimages.

## DISCUSSION

These experiments indicate that another claim for the "hypnotic transcendence of normal func-

tions" is not substantiated. *Some* individuals "hallucinate" a color and "see" its appropriate afterimage after a minimal "hypnotic induction procedure"; other individuals do essentially the same thing *without* a "hypnotic induction procedure." In fact, some *S*s do *better* on this task without the "hypnotic procedure."

However, these experiments also indicate that "trance" behavior may be an essential component in "projecting" or "hallucinating" colors. The *S*s in the supposedly "waking" experiments (Experiments 2 and 3) who stated that they had "projected" the colors, reported afterwards that they were not "thinking about anything except getting the color on the paper," they had to "really concentrate on the color," etc. As the writer has pointed out elsewhere (Barber, 1958), the term "trance" refers specifically to this type of behavior, namely, responding (overtly and symbolically) to a relatively narrow range of stimulation (or, in different terminology, "attending" to certain specific stimuli and remaining relatively "inattentive" to all other stimuli).

These experiments leave many questions unanswered. Are only *some* individuals able to "hallucinate" or "project" colors? If so, is this ability inherited, or learned, or both? Or, is this a potential ability of all human beings—possibly related to the ability to "project" images when "dreaming"—which becomes actualized only during special circumstances?

Why do individuals report negative afterimages of "hallucinated" (or "projected") colors? When an individual is "hallucinating" (or "projecting") a color is he reinstating his total original response to the color, and is the negative afterimage an innate or neurophysiological accompaniment of both the original response (to the color stimulus) and of the reinstatement of this original response (by a substitute stimulus)? There is another possibility. Since an individual has inadvertently experienced negative afterimages many times during his life, the afterimage response to the color stimulus may have become conditioned to the "verbal equivalent" of the color stimulus. In either case, further investigations of the centrifugal regulation of retinal discharge (Granit, 1955) may increase our understanding of the mechanism involved in this behavior.

## REFERENCES


BARBER, T. X. The concept of "hypnosis." *J. Psychol.*, 1958, **45**, 115–131.

ERICKSON, E. M. Critical comments on Hibler's presentation of his work on negative after-images of hypnotically induced hallucinated colors. *J. exp. Psychol.*, 1941, **29**, 164–170.

ERICKSON, M. H., & ERICKSON, E. M. The hypnotic induction of hallucinatory color vision followed by pseudo-images. *J. exp. Psychol.*, 1938, **22**, 581–588.

GRANIT, R. *Receptors and sensory perception.* New Haven: Yale Univer. Press, 1955.

ROSENTHAL, V. G., & MELE, H. The validity of hypnotically induced color hallucinations. *J. abnorm. soc. Psychol.*, 1952, **47**, 700–704.

WOLBERG, L. R. *Medical hypnosis.* Vol. 1. New York: Grune & Stratton, 1948.


Received April 7, 1958.

# SALIENCE OF NORMS AND ORDER OF QUESTIONNAIRE ITEMS: THEIR EFFECT ON RESPONSES TO THE ITEMS[1]

CHARLES Y. NAKAMURA

*University of California, Los Angeles*

IN A STUDY by Kelley (1955), it was demonstrated that differences in a person's expressed attitude on an opinion questionnaire may be directly related to the degree that a group to which he belongs is salient for him. The salience of a group was defined as "the degree to which, in a given situation, a specific group is present and prominent in a person's awareness." Kelley found that subjects who were made aware of their Catholicism through reading brief articles on the Catholic Church were more likely than control subjects to respond to a questionnaire according to Catholic norms. Charters and Newcomb (1952) reported similar results.

It would seem important to distinguish two types of norms that could be provided by the same reference group. The first, which may be called the *ideal norm*, is considered to be socially approved and ideally representative of the group, in contrast with the *behavioral norm* based on the actual behavior of group members. Although both serve as points of reference for making judgments and evaluations, the ideal norm may often differ from the behavioral.

[1] This paper was read at the WPA meetings, April, 1958.

The relative salience of these norms would appear to have implications for questionnaire construction. It is conceivable that a given question might evoke a norm which would then affect responses to succeeding questions. The effect of ordering of questions on responses to them has been discussed by several writers (Festinger & Katz, 1953, pp. 348–350; Parten, 1950, pp. 213–216). It has been shown by Ellis (1947) that the order of presentation of indirectly and directly phrased personality questions administered in a single test battery may affect test results.

The objective of this study was to investigate the effect of order of presentation on responses to different types of items on a questionnaire and to determine if such order effects can be attributed to the relative salience of ideal and behavioral norms of a given reference group as these norms are evoked by items in the questionnaire. Attitudes toward mental disorders and their treatment was selected as the subject of the questionnaire. This choice was based on the assumption that the ideal norm for a college group is an enlightened, accepting attitude toward treatment, in contrast with a behavioral norm according to which most students would acknowledge that the behavior of their fellow students is less than enlightened and that they find great difficulty in seeking counseling for their problems.

Four True-False questionnaire items (Table 1) were constructed specifically for this experiment. Items 1 and 2 serve to evoke salience of the ideal norm. It was expected that the *S*s would commit themselves to belief in this norm by answering in agreement with these statements. Item 3, the self-evaluation (S-E) item, indicates whether the *S* would be able to support his commitment to the ideal norm by responding consistently with it (False). Item 4, the group-evaluation (G-E) item, serves two purposes. It evokes salience of the behavioral norm since it refers to the behavior of other members of the group, and it gives a measure of the *S*'s opinion on the ability of others to behave consistently with the ideal norm. On the assumption that people tend to be self-protective when evaluating themselves, it was expected that when ideal norms are salient, greater item-order effects should occur in responses to items containing reference to *self* than to those that refer to *others*.

The *S*s, 116 experimental and 103 control, were students in abnormal psychology comprising men and women in their junior or senior year.[3]

In the first class meeting one of two forms (A and B) of the four-item questionnaire was distributed to each *S* in the experimental group. The two forms differed only in that in Form A, items were in the order shown in Table 1, and in Form B, Items 3 and 4 were reversed so that Item 4 preceded 3. The *S*s were instructed to read each item, in order, and indicate an answer, True or False, immediately after reading it. It was emphasized that they were not to reread any item or change an answer once they had gone on to a succeeding item. No personal identification was required. The control *S*s were administered one of two comparable forms of the questionnaire in which Items 3 and 4 were identical with those of the experimental groups', but neutral items were substituted for Items 1 and 2. The content of these neutral items were: 1. Psychology is basically a biological science. 2. Psychology is basically a social science.

TABLE 1

FOUR-ITEM QUESTIONNAIRE ON MENTAL HEALTH

1. On the basis of the present day knowledge of the causes of abnormal behavior, there should be no more stigma attached to the occurrence of a mental disorder than to the occurrence of a physical one.
2. The people in this room probably have a more enlightened view of problems of mental health than does the general public in this society.
3. Even if I were having fairly disturbing personal problems, it would be very difficult to bring myself to go to a mental health clinic to seek help.
4. Even if they were having fairly disturbing personal problems, a large majority of the members present in this room would find it very difficult to go to a mental health clinic to seek help.

In order to check on the validity of the findings of this experiment, it was repeated on 33 students in a summer session course in abnormal psychology. In this administration the *S*s were asked to place their names on the questionnaires. In the following, the different groups of *S*s are specified by the form of the questionnaire they took. In Experiment I, experimental *S*s who took Form A were designated EA and those who took Form B as EB. The analogous groups in Experiment II were designated EA′ and EB′. The control *S*s who took forms comparable to A and B were labeled CA and CB respectively. The data from the same control *S*s were used in the statistical analyses for both Experiment I and Experiment II.

## HYPOTHESES AND RESULTS

An assumption basic to the design of the experiment was that an enlightened attitude toward mental disorders constitutes the ideal norm for the reference group represented by the *S*s. This assumption was satisfactorily supported by the high

[3] The data for the controls were obtained one year after that for the experimental *S*s. However, there was no reason to believe that the two groups of students in abnormal psychology differed in any significant way.

percentage of *Ss* in both experiments who answered True to Items 1 and 2. In Experiment I, 97% of the total *Ss* answered True to Item 1 and 72% did so to Item 2. The corresponding percentages for *Ss* in Experiment II were 94% and 94%. Two specific hypotheses were tested in the study.

*Hypothesis 1. The responses to the self-evaluation item are more consistent with the ideal norm (answered False) when it follows the items that evoke this norm than when it follows the group-evaluation item that evokes the behavioral norm.* Where the S-E item follows those that evoke the ideal norm, the individual is called upon to evaluate himself under high awareness of an ideal standard to which he has committed himself. It is expected that he will tend to diminish any discrepancy between this ideal and how he really is by responding, consistently with the ideal norm, that it would not be difficult for him to seek help if he were having personal problems. In the second condition where he evaluates himself after responding to an item that evokes the behavioral norm, the heightened awareness that other people have difficulty in seeking help would make it possible for him to be more objective and admit the same of himself.

The results for Experiment I are presented in Sub-table (*a*) of Table 2. As hypothesized, the percentage of responses of False was greater for Group EA in which the S-E item followed Items 1 and 2 than in Group EB in which the S-E item followed the G-E item. The difference between proportions was significant beyond the .001 level. The corresponding results for the difference between Group EA′ and Group EB′ from Experiment II in Sub-table (*b*) show that though the samples were much smaller the findings were significant at the .02 level. In Sub-table (*c*), $p$ equals .19 for the analogous difference between the control groups, CA and CB, a finding indicative that the order effect was not significant when the items evoking the ideal norm were replaced by neutral items.

A further analysis compared each of the experimental groups, EA and EA′, with control Groups CA and CB on responses to the S-E item. Significant differences between Group EA and Group CB ($p = .01$) and between Group EA′ and Group CB ($p = .02$) were consistent with the previous results in support of the hypothesis. However, the difference between experimental Group EA and control Group CA was not significant ($p = .20$)

TABLE 2

DIFFERENCES IN RESPONSES TO THE SELF-EVALUATION ITEM (NO. 3) DEPENDING ON ORDER OF PRESENTATION IN THE QUESTIONNAIRE

| | Subjects | *N* | Order of Items | Responses % *T* | Responses % *F* | $\sigma_{D_p}$ | $z$* | $p$** |
|---|---|---|---|---|---|---|---|---|
| (*a*) | Group EA | 58 | 1, 2, 3, 4 | 40 | 60 | | | |
| | Group EB | 58 | 1, 2, 4, 3 | 76 | 24 | .092 | 3.72 | <.001 |
| (*b*) | Group EA′ | 17 | 1, 2, 3, 4 | 29 | 71 | | | |
| | Group EB′ | 16 | 1, 2, 4, 3 | 75 | 25 | .171 | 2.32 | .020 |
| (*c*) | Group CA | 51 | —, —, 3, 4 | 53 | 47 | | | |
| | Group CB | 52 | —, —, 4, 3 | 67 | 33 | .096 | 1.30 | .192 |

* The values for $z$ are corrected for continuity.
** The $p$ values are for a two-tailed test.

TABLE 3

DIFFERENCES IN RESPONSES TO THE GROUP-EVALUATION ITEM (NO. 4) DEPENDING ON ORDER OF PRESENTATION IN THE QUESTIONNAIRE

| | Subjects | *N* | Order of Items | Responses % *T* | Responses % *F* | $\sigma_{D_p}$ | $z$* | $p$** |
|---|---|---|---|---|---|---|---|---|
| (*a*) | Group EB | 58 | 1, 2, 4, 3 | 90 | 10 | | | |
| | Group EA | 58 | 1, 2, 3, 4 | 79 | 21 | .068 | 1.51 | .131 |
| (*b*) | Group EB′ | 16 | 1, 2, 4, 3 | 88 | 12 | | | |
| | Group EA′ | 17 | 1, 2, 3, 4 | 71 | 29 | .141 | 0.77 | .440 |
| (*c*) | Group CB | 52 | —, —, 4, 3 | 80 | 20 | | | |
| | Group CA | 51 | —, —, 3, 4 | 73 | 27 | .083 | 0.76 | .450 |

* The values for $z$ are corrected for continuity.
** The $p$ values are for a two-tailed test.

nor was the corresponding difference between Group EA′ and Group CA ($p = .15$). This result would indicate that part of the total order effect obtained in the responses of the experimental groups was contributed by unspecified sources other than the effect of salience of the ideal norm.

*Hypothesis 2. There is no difference in responses to the group-evaluation item whether it follows the items that evoke the ideal norm or follows the self-evaluation item.* Even when a person is highly aware of the ideal norm as evoked by Items 1 and 2, the process of evaluating others may be expected to bring into prominence the reality factor —how people really behave—and a concomitant increase in objectivity of judgment. In the order of presentation in which the G-E item follows the S-E item, it is not anticipated that the response to it will be significantly affected by the particular response given to the S-E item.

The results for the experimental groups of Experiment I and of Experiment II, and for the control groups are presented in Table 3 in Subtables (*a*), (*b*), and (*c*). The nonsignificant value of $p$ for each of the three analyses gives no grounds for rejecting the null hypothesis. However, the consistent direction of the differences suggests that the G-E item is not completely free of the item-order effect.

## DISCUSSION

The results of the study provide evidence that the distinction between ideal and behavioral norms for a given reference group may be useful for the understanding of variation in responses to questionnaire items. It was observed that the self-evaluation item was responded to differently depending upon whether it followed an item that evoked one or another of these norms, in contrast with an item that measured group-evaluation, on which item order had no significant effect. This finding is consistent with the generally accepted idea that a certain amount of self-protectiveness is evoked when people evaluate themselves on matters of importance to their self-esteem. This protectiveness was manifested, when the ideal norm was made salient, in the tendency for the *S*s to give responses that minimized the discrepancy between how they are and how they think they should be. If, however, awareness of a more realistic norm was introduced, such awareness reduced the effectiveness of the ideal norm, making possible the occurrence of more objective responses. From this point of view, the findings of this study have implications for the construction of questionnaires, inventories, interview schedules, and any instruments designed to elicit information. When the purpose is to seek a particular kind of information, the respondent may be able to give the data more candidly and completely if appropriate norms are made salient.

## REFERENCES

CHARTERS, W. W., JR., & NEWCOMB, T. M. Some attitudinal effects of experimentally increased salience of a membership group. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (Rev. ed.) New York: Holt, 1952.

ELLIS, A. A comparison of the use of direct and indirect phrasing in personality questionnaires. *Psychol. Monogr.*, 1947, **61**, No. 3 (Whole No. 284).

FESTINGER, L., & KATZ, D. (Eds.) *Research methods in the behavioral sciences.* New York: Dryden, 1953.

KELLEY, H. H. Salience of membership and resistance to change of group-anchored attitudes. *Hum. Relat.*, 1955, **8**, 275–289.

PARTEN, MILDRED. *Surveys, polls, and samples: Practical procedures.* New York: Harper, 1950.

Received April 8, 1958.

# AUTHORITARIANISM AND REACTIONS TO "SPUTNIKS"[1]

WALTER MISCHEL AND JOHN SCHOPLER[2]

*Harvard University and University of North Carolina Medical School*

THIS study investigated the relationship between authoritarianism, as measured by an F scale, and changes in predictions of potential superiority of United States versus Russia in the "space race" following the launching of "Sputnik II." Soon after the successful launching of the first man-made space satellite, "Sputnik I," on October 4, 1957, reactions to the satellite launching, including predictions about the potential space race superiority of the two major nations involved, were elicited as part of a broader project. The news that a second Soviet space satellite, Sputnik II,

[1] Thanks are due to the University of Colorado Graduate Council on Creative Research for partial support of this research, and to Marvin Nachman for his help in carrying out the study.

[2] The data for this study were collected while both authors were at the University of Colorado.

carrying a living dog, had been launched, came within a month. This event was selected as a major "natural" intervening influence, and the experimental question of the present study was then specified. Can the persons who retained their predictions of potential United States superiority in the space race, following a major objective contra-indication (Sputnik II), be differentiated, on the basis of authoritarianism as measured by an F scale, from those who changed their original predictions and conceded potential superiority to the Soviet Union? In accord with previous theorizing and research about authoritarians' tendency to resist change and to be rigid (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), it was predicted that subjects who did not change their prediction of potential United States space superiority, following Sputnik II, would be more authoritarian, i.e., would have higher F scale scores, than subjects who changed their prediction.

Research on authoritarianism as measured by the California F Scale has been reviewed and discussed by Titus and Hollander (1957) in a recent comprehensive survey. The F scale and its variations has most frequently been used to predict differences in *initial* attitudes or opinions. In contrast to this, the present study does not focus on initial differences in the particular positions of relatively high or low authoritarians but, rather, on differential *change* in opinion following exposure to new information that objectively tends to undermine the initial position. It should be clear, however, that resistance to change as investigated here is by no means necessarily synonymous with "rigidity" as defined in investigations of problem-solving in relationship to authoritarianism (e.g., Einstellung problem solutions [Brown, 1953]; Necker cube fluctuations [Jones, 1955]).

## METHOD

A total of 155 undergraduate college students, from a course in child psychology and an introductory psychology course, were administered the initial measures. Since there were minor procedural differences in administering the test materials to the two different classes, and since the use of two different samples provides independent replication, the results for each are treated separately. As a result of subject losses from absence at one or more of the three experimental sessions, the final data come from 34 *Ss* from the child psychology class (Sample A), and 70 *Ss* from the introductory psychology class (Sample B).

A measure of initial predictions (QI), a basis for measuring change in predictions (QII), and an F scale were administered to all *Ss*. The sequence of administration was as follows: QI was administered to all *Ss* on Monday, October 7, 1957, three days after the public announcement of Sputnik I; QII was administered to Sample A three days, and to Sample B eight days after the November 3, 1957 announcement that Sputnik II had been launched; an F scale was administered to Sample A two days after QII, and to Sample B one week after QII.

On QI and QII, *Ss* were asked to make the following forced-choice prediction: "Which of the two countries (U. S. or U. S. S. R.) do you think is most likely to get to the moon first?" Responses to QI were the initial measures of predicted potential superiority in the space race; responses to QII were used to assess change or lack of change from the initial predictions. *Ss* were also asked to indicate, on a five-point scale, how certain or confident they were about their prediction. These items were imbedded in a short questionnaire eliciting general reactions to the satellite launching, which included a series of three Likert-type, five-point scale items dealing with attitudes toward Russia with respect to its "people," "politics," and "culture." These questions were asked to supply control data on attitudes about Russia, to supply data for our broader study of reactions to the satellite launching, and to serve as a reasonable context in which the prediction of primary interest for the present study could be elicited.

As the measure of authoritarianism, a 30-item F scale[3] was administered to Sample B. A shorter form,[4] consisting of 10 items from the above, was administered to Sample A.

## RESULTS AND DISCUSSION

Before turning to the experimental question, it is important to note that the "influence" event, Sputnik II, was indeed effective in changing predictions about U. S. versus Soviet potential space supremacy. At the time of QI, 76% of Sample A believed the U. S. would reach the moon prior to Russia. This percentage dropped to 26% after the launching of Sputnik II. Similarly, the percentages were 68 and 33, respectively, for Sample B. In neither sample did any *S* who initially favored Russia change to the U. S. and therefore the usual tests for the significance between two proportions cannot be applied.

Our principal interest lies in the comparison of F scale scores of *Ss* who continued to predict that

[3] A 30-item scale was used consisting of the 29-item California F Scale, Forms 40 and 45, taken from Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950, Table 7, pp. 255–257), with the addition of the item listed as No. 41 on p. 250.

[4] Items 2, 6, 16, 19, 26, 29, 37, 39, 43, 44 as listed in Adorno et al. (1950, Table 7, pp. 255–257) constituted the short form.

TABLE 1
COMPARISON OF F SCALE SCORES FOR *Ss* CHANGING AND RETAINING PREDICTIONS FAVORING U. S. POST-SPUTNIK II

| Prediction: Pre- and Post-Sputnik II | Sample A | | | | | | Sample B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | Mean | *SD* | Item Mean* | *t* | *p*** | *N* | Mean | *SD* | Item Mean* | *t* | *p*** |
| U. S. Russia | 17 | 22.11 | 5.68 | 2.21 | | | 25 | 85.44 | 17.76 | 2.85 | | |
| U. S. U. S. | 9 | 29.88 | 8.33 | 2.99 | 2.83 | <.01 | 23 | 96.30 | 18.59 | 3.21 | 2.04 | <.05 |

* F scale mean divided by number of items in F scale: 10 in Sample A, 30 in Sample B.
** Single-tailed test of significance.

the U. S. would reach the moon before the Soviets with those who changed this prediction following Sputnik II. This comparison, using the *t* test for the significance of the difference between means, was made separately for the two samples and is summarized in Table 1. In both samples, the mean F scale scores of *Ss* who did not change from their initial prediction favoring the U. S. are significantly greater than for those *Ss* who changed their predictions ($p$ levels less than .01 and .05, respectively). These results are in accord with our hypothesis.

The question may be raised as to whether the relatively high and low authoritarians *changed* their predictions differentially or whether they began, i.e., made their initial predictions, with differential certainty. *Ss* were dichotomized, as closely as possible, into those giving the highest and lowest certainty ratings. Comparison of the mean F scale scores of these halves yielded no differences approaching significance in either sample ($p > .20$).

The data also permit examination of the relationship between authoritarianism and the initial, post-Sputnik I, prediction. Comparison of the means of F scale scores of *Ss* initially predicting that the U. S. will get to the moon first with those predicting that the Soviets will get to the moon first yielded no differences approaching statistical significance. (The difference in means is less than two points.) Thus, although the F scale differentiated *Ss* changing or not changing their prediction favoring the U. S. in the space race to the moon, after Sputnik II, it did not differentiate *Ss* with respect to their initial predictions.

A question of further interest deals with the relationship between attitudes toward the Soviet Union, authoritarianism, and the prediction of space superiority. The initial questionnaire (QI) provides *Ss*' reactions, on a five-point Likert-type scale, to Russian people, politics, and culture. *Ss* in both samples were dichotomized, as closely as possible, into those giving the most and least positive reactions on each of the five-point items. The mean F scale scores of these halves were then compared, using the *t* test for each of the items. None of these comparisons resulted in significant differences: in Sample A no $p$ values less than .20 were obtained; in Sample B the smallest $p$ value was greater than .10 (on the "culture" item).

These results tend to support our interpretation that the differential changes away from initial predictions favoring the U. S. are not primarily a function of differential attitudes toward the Soviet Union, e.g., differences in intensity of dislike of the Soviet Union but, rather, are a function of differential resistance to change in the face of new evidence undermining the initial position, with relatively high F scale scorers being more resistant than relatively low F scale scorers. This interpretation must remain tentative, in view of the dubious adequacy of the attitude measures (e.g., responses tended to cluster at the extreme ends of the scale items).

The present study yields no evidence on the degree of generality of the authoritarian's resistance to change in the face of new evidence that tends to undermine his initial position. Whether the relationship from here would hold for other topics, such as ones in which ethnocentrism clearly plays no role, or under different testing conditions, must remain an open question. We have demonstrated, however, that authoritarianism, as measured by an F scale, is related to resistance to change in at least one area, where influence in the direction of change stemmed from the occurrence of a natural event of major scientific and political proportions.

## SUMMARY

This study investigated the relationship between authoritarianism, as measured by an F scale, and changing or retaining predictions following a natural influence event. Following Sputnik I, college *Ss* made forced-choice predictions (U. S. versus Soviet Union) about which nation is most likely to get to the moon first. The same prediction was again elicited following Sputnik II. Among *Ss* who initially favored the U. S., those who did not change their prediction following

Sputnik II received significantly higher scores on the F scale than did those who changed their initial predictions.

## REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.

JONES, M. B. Authoritarianism and intolerance of fluctuation. *J. abnorm. soc. Psychol.*, 1955, **50**, 125–126.

TITUS, H., & HOLLANDER, D. The California F Scale in psychological research: 1950–1955. *Psychol. Bull.*, 1957, **54**, 47–64.

Received April 10, 1958.

# CONTRADICTORY RESULTS USING THE MECHOLYL TEST TO DIFFERENTIATE PROCESS AND REACTIVE SCHIZOPHRENIA

MARVIN ZUCKERMAN AND HANUS J. GROSZ[1]

*Institute of Psychiatric Research, Indiana University Medical Center*

KING (1958) recently reported two studies in which patients, classified as process or reactive schizophrenics on the basis of the case history, were shown to differ in their response to the Mecholyl test of "autonomic responsiveness." The reactive schizophrenics were found to have a greater fall in blood pressure after the administration of mecholyl than the process schizophrenics. Since data on the Mecholyl test given to 30 psychiatric patients were available here, we decided to test the findings of King in our sample.

Grosz and Miller (1958) showed that there is considerable variability in individual response to the Mecholyl test from day to day and even from arm to arm when measured simultaneously from both arms. They also pointed out the contradictory findings in the literature on this test. In this study we used the MFBP (maximum fall in systolic blood pressure), the measure used by King, to determine its reliability.

## METHOD

*Subjects.* *S*s were 30 psychiatric patients at an acute psychiatric treatment center. Of the 30 patients, 27 were diagnosed as schizophrenic, but one was excluded from the process-reactive schizophrenic comparison because her diagnosis was complicated by a convulsive disorder. Four of the schizophrenics were posttreatment. The remaining three patients were diagnosed as suffering from various neuroses. The total group of 30 was used in calculating the reliability of the MFBP.

*Procedure.* Each *S* was subjected to a mecholyl test series consisting of four separate mecholyl injections, administered into the right deltoid on four days out of seven and under the same experimental conditions. All the tests were undertaken in the morning between two to three hours after breakfast and following 45 minutes of bed rest. No patient was moved from his bed during the test period, and no *S* was moved to another ward during the test series. No somatic therapy (ECT, insulin, tranquilizing agents, or barbiturates) was applied during the test period. The only drug permitted was 1 gm. of chloral hydrate at night when necessary for sleep.

Injections were started only after stable resting systolic blood pressure readings had been obtained. The mecholyl solution was made up freshly from the powdered form for each injection. One ml., containing 10 mg. of mecholyl, was administered intramuscularly into the right deltoid. Following the injection, systolic blood pressure estimations were recorded every minute for 25 minutes. The same observer recorded all four test readings in each series with the same instrument (a mercury type sphygmomanometer). The MFBP measures for each *S* were averaged for the four trials to give the MFBP measure used in comparing the two groups.

The senior author, who had never seen the data from the Mecholyl test, was given a list of the patients' names and proceeded to classify them as process or reactive, using the criteria provided by Kantor, Wallner, and Winder (1953). Thirteen patients were classified as process and 10 as reactive. Three *S*s, presenting a mixture of features of both categories, could not be classified and were not used in the comparison.

## RESULTS AND DISCUSSION

*Reliability of the MFBP.* The intercorrelations of the MFBP scores on the four successive tests are given in Table 1. The correlations, while significant, do not reflect much stability. When the correlations are averaged, using the $z'$ conversion,

[1] Now at Albert Einstein College of Medicine, New York City, N. Y.

TABLE 1
CORRELATIONS OF MFBP ON FOUR TRIALS[a]

| Trial | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | | .47 | .50 | .62 |
| 2 | | | .61 | .63 |
| 3 | | | | .61 |
| 4 | | | | |

[a] Average $r$ = .58.

the average reliability coefficient is .58. There is a tendency for the later tests to become more stable. The fourth test correlated over .60 with each of the preceding tests. The first test appears to be the least stable, which may be due to the *S*'s anxiety, aroused by the unexpected effects of mecholyl, or by his being confronted with an unfamiliar technique.

Many persons who are quite concerned about the reliability of psychological tests seem to accept the reliability of physiological tests as a matter of faith. In speaking of the Mecholyl test, Grosz (1958) said:

> The day to day variability in the response patterns precludes the possibility of classifying patients on the basis of a single test. The marked instability of the test also contrasts with the presumably more stable characteristics with which they are being correlated. It is unlikely that correlates such as the diagnostic class, prognosis, and basic psychodynamic makeup, for example, could undergo similar fluctuations from day to day (p. 11).

One could extend this commentary to the process-reactive characterization of schizophrenia, based on the entire life history. Studies of behavioral or test correlates of the mecholyl response should be done on the same day as the test is given. If this is not done, investigators should try to obtain a stable measure from the Mecholyl test by repeating it on successive days and obtaining the central tendency of the response.

*Comparison of process and reactive schizophrenics.* The mean MFBP was 23.60 ($SD$ = 5.87) for the process schizophrenics and 16.70 ($SD$ = 6.37) for the reactive schizophrenics. The $t$ ratio was 2.58, significant beyond the .02 level. A second comparison, excluding the four posttreatment schizophrenics, yielded practically unchanged means, and the difference was still significant ($p < .05$). Since the process schizophrenics showed the greater response to mecholyl, these results are in the opposite direction to King's.

When the individual variability of the process and reactive schizophrenics was studied, using as an index of variability for each *S* the average deviation of MFBPs around his mean for the four trials, the mean A.D. was 4.33 for the process group and 3.80 for the reactive group. The difference was not significant.

The contradictory results are difficult to explain. If we had found nonsignificant results in the same direction as King's we might have suspected that slight difference in *S*s or procedures might account for the difference. However, the completely opposite nature of the findings poses questions which only further research can solve. Either both results are chance phenomena, or one is chance and one is stable, or both are real differences with some hidden but crucial variable accounting for the difference between the studies.[2]

REFERENCES

GROSZ, H. J., & MILLER, I. Two factors limiting the reliability of the mecholyl test: Variability in day-to-day responses and dissimilarity between simultaneous bilateral determinations. *J. nerv. ment. Dis.*, 1958. **127**, 417–429.

KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.

KING, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, **56**, 160–164.

Received May 12, 1958.

[2] High and low responders on the Mecholyl test were compared on Rorschach, MMPI, and Bender-Gestalt test variables. Unfortunately, the tests were not given on the same day as the Mecholyl test. No significant differences were found on the MMPI or Bender-Gestalt scores. Only 1 out of 20 Rorschach scores yielded a significant difference.

## CONCERNING THE CONCRETENESS IN SCHIZOPHRENIA

KURT GOLDSTEIN
*New York, N. Y.*

IN A PAPER published in the July 1958 issue of this Journal, Cavanaugh (1958) reported an experimental study, the purpose of which was to investigate the role of motivation in the performances of schizophrenics on concept-formation tasks. He used as means to increase the motivation the exposure of the patients to strong noises during the execution of the tests.

From his findings he concluded that schizophrenic patients are suffering from decreased

motivation. He states that his results speak against my concept of schizophrenia, assuming that I consider the concreteness of the schizophrenic to be a result of an organic defect of the brain, similar to the impairment of abstraction in organic patients. He has apparently overlooked the fact that the author has denied such an interpretation (Goldstein: 1934, 1954). He (Goldstein, 1944) stressed the "fact that in both cases there are—besides the similarities—*essential differences that should make us cautious*" in respect to such an assumption. In a previous article the author had already described the differences in the schizophrenics' concrete behavior on the tests, and pondered the possibility that the abnormal concreteness of the schizophrenics may be "considered as a way out of a conflict, as a protection against the danger of severe catastrophe."

My further studies brought me more and more to the conclusion that my conjecture was correct. Clinical observations reveal that the abnormally concrete behavior does not come to the fore under all conditions, that the same schizophrenic sometimes shows impairment of abstraction and sometimes does not, and that the change depends on the demands the patient is confronted with. I became inclined to consider the concreteness of the schizophrenics as a protective mechanism against anxiety which *originated in early youth.* (See Goldstein, 1944, p. 148.) I wrote: "The original organic unity between infant and mother is disrupted by birth. That would bring about catastrophes, anxiety, and hindrance of normal development if the disruption is not repaired, if a new unity between infant and 'world' is not established in this stage. To what extent that can be achieved depends on conditions of the environment, particularly the behavior of the mother or other persons in the environment." If it is not perfectly accomplished, catastrophes occur which the infant is not able to bear. He reacts to them with withdrawal, the only protective mechanism which can be built in this stage of infancy when the abstract attitude has not yet developed.

Persons who later become schizophrenic retain the habit of reacting to dangerous situations with abnormal concreteness. They react so particularly to persons because the latter easily produce anxiety as an aftereffect of experiences in infancy. The abnormal concreteness of the schizophrenic appears thus as a *secondary* phenomenon; it is *not the effect of an organic defect*; it does not represent a damage of abstraction, a deterioration of the mind, or a defect in thinking. It is an *expression of the restriction in the use of the highest mental capacity.*

When one wants to understand the behavior of schizophrenics, one must be aware that their concrete behavior shows definite deviations from that of normals. Normal people also surrender under certain conditions to activities of an unreflected character. They are induced to behave concretely in special situations. But this does not occur passively; some intention is always involved. They are not forced to do it because their activities are always under some control by the abstract attitude. The patient is not only forced to react to a strong stimulus, an object or situation; his reactions occur without this control and show therefore the effect of "isolation" (Goldstein, 1934, p. 133). A great number of schizophrenic reactions become understandable when considered from this point of view.

I think the results of the experiments of Cavanaugh can be interpreted from our viewpoint regarding the concreteness of the schizophrenics. The author reports that under the irritative noise condition (produced by disturbing but not "injurious" stimulation), "the results in concept-formation were better. The schizophrenics approximated the performance level of the normals" (Cavanaugh, 1958, p. 11). He assumes that the improvement was due to "an experimentally induced increase in schizophrenic motivation." If I understand him correctly, he considers the schizophrenic to be in a defensive condition, which it is important for him to keep in order to avoid dangerous situations. The reaction to the strong noise, one could assume with the author, distracts the patient from the maintenance of this defense, so he is able to act more "normally" and therefore behaves more correctly in the concept-formation tests. One can also assume that concrete behavior may need less initiative than abstract behavior, because it is directly determined by the stimulus. When the power of motivation is diminished in schizophrenia, the experimental improvement due to noise then may be understandable as an effect of increased motivation.

But to understand the result of the experiment, is it necessary to assume that the schizophrenic is suffering from a diminution of motivation in general? That would not correspond to the observations. The schizophrenic often seems to act even under very strong motivation, for instance, in respect to the maintenance of the concreteness, at least in special situations which are of vital interest. Indeed, I would not consider that a correct description of the situation. The patient does not want to be concrete, he does not know that he is concrete, he reacts so that he can come to terms with the task, as all normal persons would, and applies the infancy-developed habit, i.e., he reacts in the concrete way. Thus, one could at least not say the "motivation" is reduced in general.

When one wants to characterize the behavior of

the schizophrenic by applying the terminology of motivation, one can only say that the schizophrenic appears sometimes motivated, sometimes not, but an explanation of when he is and when he is not can be given only if one considers the present condition carefully from the point of view of whether or not it is apt to produce anxiety. If he tries to handle a dangerous situation by concreteness and is successful, then he will appear strongly motivated; if he is not allowed to do that or is made aware that the result does not correspond to the task, he will not appear to be motivated and will appear more or less totally withdrawn. When the situation is not dangerous, he will behave concretely *or* abstractly, depending on the greater successfulness of the one or the other procedure.

Thus, I think the simple statement that the patient's motivation is decreased does not do justice to the facts. Concerning the explanation of the experiment of Cavanaugh, I would say the particular strong auditory stimulus diminishes reactions in other parts of the organism. This is in accordance with the way the normal organism reacts to strong stimulations. The reaction to the strong stimulation does not let the anxiety come to the fore, which belongs to the reaction in another part of the organism. In the experiment the habitual tendency to be concrete in a threatening personal situation would therefore be diminished or even hindered, and the patient will use his abstraction, by which alone the test can be correctly performed. From this it is understandable that the more normal reactions to the test last no longer than the irritation by the stimulus, as the author reports.

Thus, the improvement of the behavior in the experiment can be considered as the result of the elimination of the effect produced by the specific form of anxiety in schizophrenia. The test situation produces anxiety, as all test situations easily do, because it is strange and is experienced in relation to the demands of *our* world, the personal world which is so dangerous for the patient. From this it becomes further understandable why the concreteness does not show in respect to all test situations in an equal way, which we had occasion to observe because the test materials we used were different. The difference consists probably in the different experiences of danger related to the various materials directly or to the schizophrenic's associations to them. We can judge a performance of the patient only when we analyze what it represents in respect to the possibilities of the total personality of the patient and of the use or non-use of them in the particular situation. Thus, the results of the tests may show a difference according to the degree of the development of a protecting transference, a point which is not always given enough attention in psychological experiments.

I hope that this discussion may eliminate forever the wrong assumption that I consider the abnormal concreteness in schizophrenics as an effect of an organic defect which damages the capacity of abstraction and, moreover, will help to clarify the complexity of the problem we are confronted with in the abnormal concreteness of the schizophrenic and its different appearance in different situations.

## REFERENCES

CAVANAUGH, D. K. Improvement of schizophrenics on concept formation tasks as a function of motivational change. *J. abnorm. soc. Psychol.*, 1958, **57**, 8–12.

GOLDSTEIN, K. *The organism.* New York: American Book, 1934.

GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorders. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944.

GOLDSTEIN, K. Transference in treatment of organic and functional nervous disorders. *Acta Psychotherapeut.*, 1954, **2**, 334–353.

Received December 17, 1958.

# PATTERNS OF PARENTHOOD AS THEORIES OF LEARNING[1]

JANE LOEVINGER

*Jewish Hospital of St. Louis*

GERHART PIERS (in a talk which I quote from memory) has divided methods of learning into three types: learning by reinforcement, by insight, and by identification. All three types of learning unquestionably occur, and theories of learning espoused by professional psychologists must account for those facts, whatever they take as the prototype of all learning.

Any consistent method of child-rearing contains

[1] Preparation of this note was supported by Research Grant M-1213 from the National Institute of Mental Health, Public Health Service.

by implication a theory of how children function, particularly how they learn. One can easily set up a correspondence between well-known patterns of parenthood and the three types of learning. Corresponding to any pattern of child-rearing there is, then, a "theory of learning," emphasizing one type of learning at the expense of others. Theories of learning held by parents are, of course, far more naive and uncomplicated than similar theories held by psychologists. To avoid confusion, the term "parental theory" may be used to distinguish the implicit learning theory.

The disciplinarian parent apparently believes that any wrong thing a child does will be continued indefinitely if the parent does not see that it is punished. While psychological research has tended to emphasize rewards as more effective than punishments as reinforcing agents, disciplinarian parents emphasize punishments as reinforcers.

Apparently insight learning is assumed to predominate by those parents, once reputed to be numerous, who believe that every demand made on a child must be rationalized and explained.

Finally, the typical permissive parent must surely believe that the socialization of his child takes place by means of the child identifying himself with the well-socialized parent.

Consider the following situation. Five-year-old Johnny is beating on his two-year-old sister Sue. Mother comes in. Let us assume that every mother will want to prevent repetition of such behavior. What will she do?

Mother One believes that if Johnny does not feel pain, he will repeat the behavior at every coincidence of impulse and opportunity. She therefore punishes him sharply, thus demonstrating her adherence to a parental reinforcement theory of learning.

Mother Two believes that Johnny can be shown how wrong his conduct is and sets about to persuade him. She believes in a parental insight theory of learning.

Mother Three believes that Johnny wants to grow up to be like his parents. If she punishes him harshly, he will learn that it is all right for the bigger one to be mean to the littler one if he or she feels like it; so his behavior is less likely to be repeated if reprimanded gently than if dealt with harshly. She believes in a parental identification theory of learning.

There is one fallacy common to all parental learning theories. Kelly (1955) points out that we are not victims of our history but only of our construction of that history. Kelly finds in that fact hope for the psychotherapist. But just as it gives hope for the therapist, it generates despair for the parent. A parent can decide to beat his child, but he cannot decide how the child will construe the beating. Nor, if he abstains from punishing, can he decide how the child will construe the abstention.

Rules for rearing children are beyond the scope of this note, indeed, beyond the competence of the writer. But one superordinate rule can safely be stated: Whatever the parent's theory of learning, the child will in fact be learning by an alternative method. Thus the son of Mother One is probably identifying with a punitive, disciplinarian adult; for the son of Mother Two it is being stamped in that beating on sister has no painful consequences; while the son of Mother Three has probably discerned, "Aha! I can get away with it." The explanation of why a child shifts his mode of learning to escape his parent's vigilant efforts at socialization is not difficult. He is attempting to defend the gratification of his impulses, and in this respect he is not altogether different from his parents.

The foregoing formulation helps to solve two riddles. Why is the battle between the generations fought, generation after generation, with such vigor? And why is it that experts on child-rearing are not conspicuously more successful at the art than those less expert?

The failure of expertise in child-rearing was foreshadowed in 1909 with Freud's (1925) publication of *Analysis of a Phobia in a Five-Year-Old Boy*, for little Hans was the child of two of Freud's followers. One should not make too much of the fact. He was not necessarily the most neurotic child in Vienna, merely the one that Freud had opportunity to observe and indirectly to treat. Nonetheless, the occurrence of so severe a phobia in the child was a striking omen.

Reasons have been advanced for the failure of children of experts to be vastly superior to others in their adjustment. Without disputing or discounting those reasons, one can focus on a slightly different one. The experts know what other parents did wrong, and they avoid those errors. But while they avoid the errors of parents in other houses, their children contrive to defend their instinctual gratification against the parents in their own house. In current terms, a shift in parentmanship is countered by a shift in childmanship.

The battle between the generations is commonly accounted for by the fact that parents have need to socialize their children, and the children forever battle against the socializing process. This view is the one being elaborated here. But it is not quite the whole story. A useful way to test a theory is to see what happens in the most extreme cases. Redl and Wineman (1957) have depicted extreme cases of "children who hate." Many of the sentiments of those children, such as "Grownups don't

want kids to have any fun," are echoed occasionally in almost all homes. But the ferocity and implacability of the war with adults is entirely disproportionate to what takes place in an ordinary household. Were their parents, then, so rigorous in their attempts at socialization? On the contrary, the parents of those children presented a picture of impulsivity no less striking than that of the children. The abuses to which the children were subjected could hardly be called punishments; they did not appear to result from any theory of how children learn but rather were crude lashing out on impulse. The picture of parent-child relations in *The Aggressive Child* is a conspicuously undesirable one, both prima facie and in terms of outcome. It serves to demonstrate that not all parents are informed by a parental theory. The battle between the generations is never more vicious than when all pretense of representing the interests of society is dropped and it becomes the parent's impulsivity versus the child's.

A general theory of the battle between generations must account for all of the cases. It must therefore read that the child's impulse gratification conflicts with the needs of society, represented by parents, to socialize him, as well as with the parent's own impulse gratification. The normal parent, to be sure, satisfies many of his desires in and through his children. But moment by moment and day by day the needs which the children gratify are not always uppermost. The presence of an infant or child in the household necessarily imposes delay or surrender on many of the parents' wishes.

The conclusions of this discussion can be stated simply, though they do not exactly simplify life. Every consistent pattern of child-rearing embodies a theory of learning, and all those parental theories are substantially wrong. However, any parental theory is better than none.

Is it possible to base one's pattern of child-rearing on a more nearly realistic theory of learning? That is an intriguing question. In view of the adaptability of the normal child in shifting his tactics to match those of his parents, such a method would require constant reconsideration and change. Yet inconsistency, so the child-rearing experts tell us, is one of the worst faces a parent can turn to his child. Possibly, however, inconsistency got its bad name not from conscientious parents trying to outwit their children but from the label being applied to such parents as Redl and Wineman have sketched.

If, as the present discussion suggests, parental theories are more wrong than right, how does it happen that it is better to have one than not? The chief value of a parental learning theory may well be in providing a model for the child of curbing one's own impulses out of regard for the future welfare of another. The very oversimplification of parental theories may serve to make accessible to the child that his parent is acting on principle rather than on impulse. To say this is to lay emphasis on learning by identification. But probably most psychologists, whatever their professional theories, act in relation to their own children as if they expect them to learn chiefly by identification.

"All I say is by way of discourse, and nothing by way of advice. . . . I should not speak so boldly, if it were my due to be believed" (de Montaigne, 1913, p. 283).

## REFERENCES

DE MONTAIGNE, M. *The essays of Michel de Montaigne.* Vol. III. C. Cotton (Trans.); W. C. Hazlitt (Ed.) London: G. Bell, 1913.

FREUD, S. Analysis of a phobia in a five-year-old boy. In *Collected papers.* Vol. III. London: Hogarth, 1925. Pp. 149–289.

KELLY, G. A. *The psychology of personal constructs. Vol. II. Clinical diagnosis and psychotherapy.* New York: Norton, 1955.

REDL, F., & WINEMAN, D. *The aggressive child.* Glencoe, Ill.: Free Press, 1957.

Received February 16, 1959. Prior publication.

# PSYCHOLOGICAL MONOGRAPHS

## General and Applied

Collector's items of the future!*

*A Dual Approach to Rorschach Validation*, by J. D. Palmer. No. 325. *$1.00*

*A Prognostic Scale for Shock Therapy*, by M. J. Feldman. No. 327. *$1.00*

*The Organization of Attitudes Toward the Negro as a Function of Educ...*, by L. A. Kahn. No. 330. *$1.50*

*Studies in Reputation: I. Sex and Grade Differences in School Children's Evaluations of Their Peers. II. The Diagnosis of Social Adjustment*, by R. D. Tuddenham. No. 333. *$1.50*

*Personal Security as Related to Station in Life*, by A. L. Knutson. No. 336. *$1.00*

*The Performance of Schizophrenic and Normal Individuals Following Frustration*, by Harold Wilensky. No. 344. *$1.00*

*Parents' Attitudes vs. Adolescent Hostility in the Determination of Adolescent Sociopolitical Attitudes*, by Kenneth Helfant. No. 345. *$1.00*

*Personality of the Convulsive Patient in Military Service*, by T. W. Richards. No. 346. *$1.00*

*Judging Interests from Expressive Behavior*, by N. L. Gage. No. 350. *$1.00*

*The Effects of Analeptic Drugs in Relieving Fatigue*, by R. H. Seashore and A. C. Ivy. No. 365. *$1.00*

(*Random selection from Volumes 65, 66, and 67. Write for complete list 1951–1958.)

If any of these monographs are missing from your library—lost, strayed, or stolen—order now, before they go out of print, from

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N.W.
Washington 6, D. C.

# PSYCHOLOGY AND REHABILITATION

**Beatrice A. Wright, Ph.D.**
*Editor*

Based on the proceedings of an Institute on the roles of psychology and psychologists in rehabilitation, sponsored by the American Psychological Association and supported by a grant from the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

The volume examines the principles and assumptions underlying the modern philosophy of rehabilitation, paying special attention to the roles and functions of psychologists as members of an interdisciplinary team. Problems of training are explored in detail and in the context of the changing nature of the field. Areas of needed research are highlighted.

Price $1.50

Order from

**American Psychological Association, Inc.**
Dept. PAR
1333 Sixteenth Street, N.W.
Washington 6, D.C.